Critics assail scientist behind dramatic superconductivity claims p. 19

Discharging radioactive wastewater into the ocean p. 31

Qubit platform built atom by atom p. 87

Science

$15
6 OCTOBER 2023
SPECIAL ISSUE
science.org
AAAS



VISIT…

LANZAROTE Caliente.com

GLIMPSES OF GENOMES PAST



By **Corinne Simonti** and **Madeleine Seale**

Ancient DNA has radically altered our understanding of the evolutionary history of many plants and animals, as well as our own. It has revealed areas of our genomes that were inherited from closely related archaic hominins, granted insight into past ecosystems, and unveiled extinct species and populations from across the tree of life. Each sample of ancient DNA is a snapshot of a place and time, giving a brief glimpse into the past. A better understanding of the DNA that was inherited by modern individuals—and, sometimes more importantly, the DNA that wasn't—can inform us about many evolutionary processes, including adaptation, speciation, and domestication.

As the field of ancient DNA matures, we now have the foundational understanding to appreciate the finer details granted by these snapshots. Studies of past plants and animals can help to illuminate the changing ecological landscape, as well as our role in those changes. When human studies are done ethically and responsibly, we can integrate with communities and rediscover stories that have been lost or erased from history, thereby possibly reconnecting modern people to historical relatives. Beyond these focused investigations, however, the technical limits of ancient DNA remain unclear. Although methodologies for capturing ancient DNA have greatly improved, how long DNA can survive environmental damage in different climates and conditions is still an open question.

Though we will never be able to get a full picture of the past, studies of ancient DNA continue to reveal previously unknown aspects of history across organisms. This special issue examines the changing landscape of how ancient DNA is studied today, including previously untapped sources, improvements in technology, and ethical challenges, and what we've learned about ourselves through ancient DNA.

10.1126/science.adk9651

INSIDE

REVIEWS

Deep-time paleogenomics and the limits of DNA survival p. 48

Going local with ancient DNA: A review of human histories from regional perspectives p. 53

Ancient DNA genomics and the renaissance of herbaria p. 59

RELATED ITEMS

NEWS STORY p. 24

PODCAST



Herbaria are a growing resource for ancient DNA studies. The plant specimens shown here are *Ambrosia artemisiifolia* (this page), also known as common ragweed, and *Solanum tuberosum* (opposite page), also known as potato. Herbarium specimens contain not only biological materials but also notes from their original collection.

PHOTO: ROYAL BOTANIC GARDENS KEW

REVIEW

Going local with ancient DNA: A review of human histories from regional perspectives

María C. Ávila-Arcos[1]*†, Maanasa Raghavan[2]*†, Carina Schlebusch[3,4,5]*†

Ancient DNA (aDNA) has added a wealth of information about our species' history, including insights on genetic origins, migrations and gene flow, genetic admixture, and health and disease. Much early work has focused on continental-level questions, leaving many regional questions, especially those relevant to the Global South, comparatively underexplored. A few success stories of aDNA studies from smaller laboratories involve more local aspects of human histories and health in the Americas, Africa, Asia, and Oceania. In this Review, we cover some of these contributions by synthesizing finer-scale questions of importance to the archaeogenetics field, as well as to Indigenous and Descendant communities. We further highlight the potential of aDNA to uncover past histories in regions where colonialism has neglected the oral histories of oppressed peoples.

Exponential increases in ancient human genomic data over the past 10 years has enabled the reconstruction of past migrations, admixture between human populations and between anatomically modern human lineages and other hominins, natural selection, and infectious diseases [e.g., (*1*–*3*)]. Most of this wealth of information has been generated by a few groups of investigators whose primary focus, until recently, has been large-scale events such as the out-of and back-to Africa migrations and the initial peopling of continents, with an overrepresentation of European-focused studies. Since then, there have been calls to diversify paleogenomic research, to consider more local questions, and to go "beyond broad strokes" to sociocultural aspects such as kinship, marriage practices, and how polities and power were structured in the past (*4*). Although many big-picture migration questions remain unanswered, particularly in non-European contexts [see Boxes 1 to 3 for a summary of existing broad-scale ancient DNA (aDNA) research], there is growing literature focusing on site- and region-specific scales. It is noteworthy that in several broad-level studies, there are cases, at times deemed genetic "outliers," that do not conform to the big migration narratives and/or regions that have a different demographic history compared with the reported continental-scale pattern. Such cases offer exciting and realistic snapshots of processes occurring at local scales. In this Review, we draw upon select paleogenomic studies (Fig. 1) that illuminate local histories in the Americas, Asia, Oceania, and Africa, and discuss the potential of the field to add fine-scale resolution to human demography, social structures, and health over time.

[1]International Laboratory for Human Genome Research, Universidad Nacional Autónoma de México, Querétaro, Mexico. [2]Department of Human Genetics, University of Chicago, Chicago, IL, USA. [3]Human Evolution, Department of Organismal Biology, Uppsala University, Uppsala, Sweden. [4]Palaeo-Research Institute, University of Johannesburg, Johannesburg, South Africa. [5]SciLifeLab, Uppsala, Sweden.
*Corresponding author. Email: mavila@liigh.unam.mx (M.C.A.-A.); mraghavan@uchicago.edu (M.R.); carina.schlebusch@ebc.uu.se (C.S.)
†All authors contributed equally to this work.

Connecting ancient and modern peoples

There is a recurrent interest in identifying genetic connections between ancient and present-day populations. Although there can be certain technical limitations in assessing population continuity with absolute certainty, including a bias in the availability of present-day reference populations, the identification of relationships at different levels of resolution between ancient and modern peoples has generated rich knowledge in human paleogenomics. For instance, it has allowed us to learn when the formation of the genetic structure observed today was established, the characteristics of different past demographic events such as gene flow from other known or "ghost" populations, and to prove or contest hypotheses of population replacement. These studies have even allowed the identification of ancestors and reaffirmed the connection to the land for Indigenous populations.

Paleogenomic studies targeting African populations have assessed the existence of connections between African and Eurasian and Middle Eastern populations. For example, the use of aDNA has given insights about the origin of the genetic connection of present-day North Africans with Eurasian, Middle Eastern, and, to a lesser extent, sub-Saharan populations. The origin of this genetic makeup had been a subject of debate, with some attributing it to Paleolithic back-to-Africa migrations and others to migrations during the Neolithic period. Studies analyzing aDNA from 15,000-year-old remains in current-day Morocco have revealed substantial Eurasian genetic contributions that predate the Holocene and the emergence of farming practices (*5*). The Eurasian component of these Epipaleolithic individuals was related to Levantine Natufian populations, whereas a separate study (*6*) showed genetic continuity between Epipaleolithic hunter-gatherers and Early Neolithic farmers in Morocco, indicating a cultural transition rather than population replacement. Late Neolithic individuals exhibited an Iberian component (*6*). These findings suggest that the spread of farming practices into this region of North Africa involved the movement of both ideas and people.

In West Africa, an aDNA study on four individuals between 8000 and 3000 years old from a rock shelter at Shum Laka, Cameroon (*7*), challenged the assumption that the cave displayed the cultural development sequence of Bantu speakers from hunter-gatherers to more settled societies before the onset of the Bantu expansion. The genetic analysis revealed a surprising association of the ancient individuals with rainforest hunter-gatherers instead of Bantu-speaking groups. Although this unexpected finding questions the in situ cultural sequence of the Bantu-speaker culture at Shum Laka, it is still possible that Bantu-speaker ancestors coexisted with rainforest hunter-gatherer ancestors who were buried in the caves.

In Asia, the search for genetic connections between ancient individuals and present-day populations has yielded knowledge on the local human genetic histories of regions such as Korea and Tibet. A study by Gelabert *et al.* (*8*) investigated the pre- and protohistory of Korea by shotgun sequencing eight ancient individuals from the Korean Three Kingdoms period (57 to 668 BCE) that precedes the Korea Kingdom and the present-day national identity of Korea. The authors found that present-day Koreans are closely related to the Three Kingdoms period individuals and detected a past population structure that may be indicative of diverse groups coinhabiting the region. Similarly, paleogenomic studies of ancient individuals from the Tibetan Plateau and the Himalayan arc (*9*, *10*) have shed light on early population structure and genetic diversification of plateau-specific lineages, and have also revealed ancient gene flow between this region and East, South, and Central Asia, which has shaped its present-day genetic landscape.

Similarly, paleogenomics projects performed between researchers in partnership with Indigenous communities in North America have revealed the relationships between Ancestors and Indigenous descendants, as well as a more detailed account of the regional demographic history (*11*–*13*). By studying genome-wide data from ancient individuals from the Pacific Northwest Coast in Alaska, a region of interest because of its association to the coastal dispersal hypothesis (see Box 3), Lindo *et al.* (*11*) identified an early ancestral genetic lineage seemingly stemming from a late Pleistocene coastal migration into the Americas. This ancestral lineage, discovered from the low-depth genome sequencing of an ~10,300-year-old

individual known as "Shuká Káa" ("Man Ahead of Us"), was different from the ones previously identified from the late Pleistocene (*14*) or early Holocene [Kennewick Man, referred to as "The Ancient One" by Native Americans (*15*)]. The authors also found that population structure existed among Native North American groups at least since the late Pleistocene. The same authors conducted a study of the genomes of eight present-day tribal members and 12 ancient individuals from two archaeological sites on the Muwekma Ohlone tribe in the San Francisco Bay land spanning ~2000 years (*13*). The analyses of these present-day and ancient genomic data confirmed genetic continuity in the

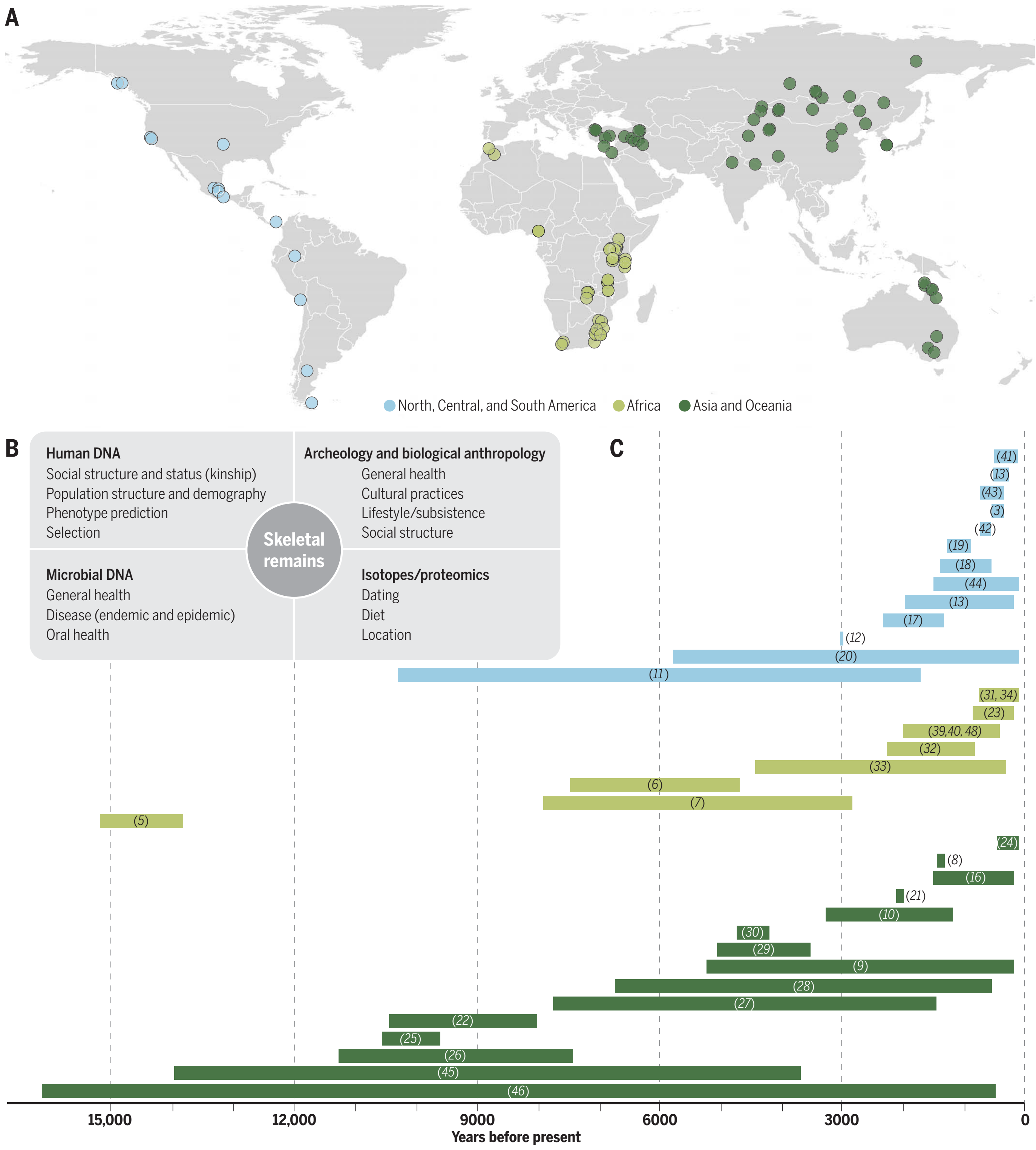


Fig. 1. A non-Eurocentric view of human paleogenomic studies. (**A**) Map showing paleogenomic studies with a focus on non-European regional questions highlighted in this Review. Average geographic coordinates are shown for studies with samples from several proximal locations. (**B**) The information that can be obtained from skeletal remains to reconstruct past histories can be divided into different categories. (**C**) Age range of the individuals sampled in the studies highlighted in this Review.

San Francisco Bay Area for the past 2000 years. These results stood in contrast to previous reconstructions based on archaeological and linguistic information, adding knowledge of Indigenous population history in California and western North America more broadly.

Finding genetic connections between ancient human remains and present-day populations is of particular relevance to Indigenous communities, who bear strong links to their land and Ancestors and may seek repatriation of Ancestral remains. For example, a study from Australia sequenced 10 nuclear and 27 mitochondrial genomes from precontact Aboriginal Australians and revealed past population structure, including an east-west genetic gradient also seen today, as well as signals of regional genetic continuity (*16*). In the Americas, finding genetic continuity has been the form rather than the exception, as it has been detected in the Pacific Northwest (*11*, *12*), California (*13*), Mexico (*17*), Central America (*18*), and Patagonia (*19*, *20*) as early as 10,000 years ago and as late as 1000 years ago.

Sociocultural revelations

A few recent works have used paleogenomics to investigate local aspects of population histories and societal structures at the community level, including past sociopolitical dynamics through investigations of burial traditions, kinship, sex, and social status. We highlight a few studies on this theme below, focusing on non-European contexts, and direct readers to Racimo *et al.* for deeper coverage of European case studies on sociocultural aspects (*4*).

Lee *et al.* (*21*) focused on the Xiongnu empire, which established the first "nomadic imperial" system in the Eastern Eurasian steppe. From archaeological and historical evidence, it is suggested that nomadic and sedentary groups converged into this region between 200 BCE and 100 CE, resulting in a multiethnic society. The authors investigated how this culturally characterized diversity, human movements, and ethnopolitical organization at the local scale translated to genetic diversity. Through the analysis of 18 ancient individuals from one aristocratic and one local elite cemetery, they found that both high- and low-status burials harbored high genetic diversity, likely representing the overall diversity of the Xiongnu empire. Lower-status individuals, often located in the periphery of the higher-status or elite burials, harbored higher genetic heterogeneity compared with the higher-status individuals, suggesting that the former were drawn from diverse parts of the empire. The authors also detected extended family groups and found that the highest-status individuals were primarily women, suggesting that the Xiongnu women likely had prominent sociopolitical roles.

A similar study led by a Turkish team explored how early sedentary Anatolian Neolithic societies were organized (*22*). By inferring kinship and pedigrees for 59 Neolithic Anatolian individuals, including 22 ancient genomes generated from two important sites from this period, the authors observed a changing trend over time with respect to burial and kinship practices. In the 9th to 8th millennia BCE, related individuals tended to be buried together in domestic structures, whereas in the 7th millennium BCE, burial structures included biologically unrelated individuals, indicating a shift in the societal structure with respect to how familial relationships were regarded.

In Africa, a recent study of the Swahili coast indicated asymmetric social interactions and sex-biased admixture between groups during back-to-Africa gene flow from various traders introducing different Eurasian components at different time periods (*23*). The study showed substantial genetic mixing between African and Asian populations, particularly those from Persia and India, between 1250 and 1800 CE, coinciding with the spread of Islam in the region. Subsequent interactions with Asian and African groups have further changed the ancestry of present-day Swahili coastal populations.

In Australia, using uniparental markers from five Ancestral Aboriginal Australian remains from Cape York in northeast Queensland, a study found that individuals buried both in beach and rock shelter contexts shared close genetic affinities with present-day communities from the same region (*24*). By contrast, an individual with a more distinctive beach burial was genetically closer to present-day populations from another region, specifically central Queensland or New South Wales, thereby gaining more resolution on past mortuary practices.

As paleogenomics turns to more site-specific questions, aDNA in tandem with archaeology can aid in the reconstruction of past societal facets. Such studies can bring us closer to unraveling complexities of human societies that are not always captured by larger narratives of migrations and admixture. However, we do need to be cognizant of ascribing identities to ancient individuals and societies that may be biased by present-day worldviews and practices.

Subsistence and lifestyle

The integration of paleogenomic studies, isotopic analyses, and archaeological inferences has yielded substantial contributions to our understanding of the evolution of subsistence strategies. By providing valuable insights into the genetic characteristics of hunter-gatherers, agriculturists, pastoralists, and marine-based societies over time, these interdisciplinary approaches have deepened our knowledge of how subsistence practices change over time.

Genetic diversification of populations with different lifestyles and subsistence strategies is a particular recurrent theme in paleogenomic studies in the Americas (*17*, *19*, *20*). For example, the analysis of the genome of a 3000-year-old woman referred to as "Tatóok yík s sháawat" found in southeast Alaska revealed a divergence time of at least 6000 years ago between coastal and inland populations (*12*). At the other end of the Americas, studies on the genomes of ancient Patagonians indicated an initial separation between maritime and terrestrial hunter-gatherer populations after the formation of the Magellan Strait ~8000 years ago, followed by diversification among marine hunter-gatherers. Apart from providing an approximate date for the appearance of marine hunter-gathering as a subsistence strategy, the study of these ancient Patagonian genomes also refined the timing of cultural transitions in the region (*19*, *20*).

In Asia, paleogenetic studies suggest regional heterogeneity in human expansions associated with the spread of agriculture and domestication. aDNA from Upper Mesopotamia, home to some of the earliest sedentary societies leading up to the Neolithic Transition, suggests genetic connections between this region and Neolithic Anatolia that shaped the genetic signature of early Anatolian farmers

Box 1. Contributions of aDNA to the study of human evolutionary histories in Africa.

The concentration of landmass on the equator creates hot and humid conditions in Africa, which has largely hindered the preservation of aDNA. This, coupled with financial constraints in developing countries, has delayed aDNA research in Africa and placed it in its early stages compared with other continents. Thus far, research in Africa has primarily focused on deep-time population structure (*5*, *7*, *32*, *48*) and large-scale Holocene migration patterns connected to herding or farming practices and trading (*1*). aDNA studies using synchronous genetic analysis of modern and ancient African populations have shown that the earliest population divergence among modern humans in Africa occurred between the Khoe-San groups of southern Africa and other African populations, whereas the hunter-gatherer groups of the central African rainforest diverged later (*32*, *48*). A recent study on aDNA from Shum Laka (Cameroon) in West Africa proposed a radiation model involving lineages leading to Khoe-San hunter-gatherers, rainforest hunter-gatherers, East and West Africans, and a "ghost modern" lineage that no longer exists (*7*). A "ghost-archaic" source is also included in the model, and West African groups exhibit low levels of "ghost archaic" admixture, together with a larger amount of ghost-modern admixture. aDNA from anatomically premodern human remains might be crucial to resolving the question of archaic admixture in Africa, as it did for Neanderthal admixture outside of the continent.

(*25*, *26*). Similarly, early millet-farming societies in the Yellow River and West Liao River basins in northern China showed substantial shifts in their genetic profiles over the past six millennia that paralleled subsistence transitions (*27*). Specifically, intensification of millet farming in the West Liao River region was associated with an increased Yellow River genetic affinity. The role of human migrations in the introduction of dairy pastoralism in Central Asia has been investigated by several aDNA studies. They have, for example, found associations between this subsistence and Afanasievo-related gene flow into Mongolia (*28*), but no such association was detected farther west in Xinjiang, where the earliest Tarim mummies likely represent a local agropastoralist population (*29*). In South Asia, genomic analysis of a single individual affiliated with the Bronze Age Indus Valley Civilization and 11 genetic outliers from Bronze sites in Iran and Turkmenistan can be modeled as a key ancestral source for many present-day South Asians (*30*). Lack of Anatolian farmer-related genetic ancestry in these ancient individuals could suggest local agricultural development in South Asia.

Paleogenomic studies in Africa have been focused on the impact of agriculture on population distributions. For example, there have been studies of the Bantu expansion, which started ~4000 years ago and involved West African groups who spread throughout sub-Saharan Africa, replacing or intermixing with existing groups (*31*), or that of herder communities from Northeast Africa, who also migrated southward, settling in East Africa and even reaching southern Africa (*32*, *33*). These studies have used aDNA to deepen our understanding of population movements and the interaction between different groups during this transformative period. Several aDNA studies in Africa also provide comprehensive information on archaeology, social structure, subsistence practices, and inferred diets [e.g., (*7*, *33*)]. As an illustration, in a study focusing on individuals from central south Africa, isotope analysis indicated a low-trophic diet associated with farming, which contradicted the previous hypothesis of Stone Age hunter-gatherer affiliation (*34*). Subsequent aDNA studies confirmed their identification as Iron Age farmers genetically connected to present-day Bantu-speaking populations in the surrounding regions (*31*).

Integrating paleogenomic studies, isotopic analyses, and archaeological inferences has expanded our knowledge of subsistence strategies and associated population dynamics, holding promising prospects for understanding the complexities of human history and cultural interactions in the future. Additionally, we acknowledge a rich body of literature on animal and plant domestication (*35*) and bone assemblages from archaeological middens (*36*, *37*) that add to our understanding of past diets and subsistence.

Disease and health

aDNA studies have substantially advanced our understanding of local patterns in the disease and health of ancient populations (*38*). By analyzing disease dynamics, transmission patterns, and genetic adaptations, these studies provide valuable insights into the complex interplay among lifestyles, past human migrations, and diseases in different regions (*35*).

In Africa, for example, a paleopathology study of the skeletal remains of a 2000-year-old hunter-gatherer boy related to modern-day Southern African Khoe-San groups indicated poor health, leading to his death at a young age estimated at ~7 years (*39*). The shotgun metagenomics analysis revealed that the boy was infected with *Rickettsia felis*, a flea-borne pathogen that causes a typhus-like disease (*40*). The finding that *R. felis* existed at least 2000 years ago among southern African Stone Age hunter-gatherer societies challenged previous assumptions that *R. felis* was a new or emerging pathogen associated with sedentary lifestyles.

In the Americas, the study of pathogen aDNA has been useful in understanding the effects of European colonization on disease prevalence [e.g., (*41*)] and the extent of pathogens that were present before contact in the Americas [e.g., (*42*, *43*)]. For instance, aDNA extracted from precontact individuals from Peru with tuberculosis-associated lesions revealed that the strain responsible for these infections likely arrived in the Pacific Coast of South America through marine mammals (*43*). Similarly, comparison of oral pathogen aDNA between the precontact and postcontact periods has shown a replacement of Indigenous strains by European strains, as well as a higher prevalence of these pathogens in postcontact times (*42*, *44*). Several pathogens have been detected through aDNA in the Colonial era, including the infectious agents responsible for paratyphoid fever, syphilis, and hepatitis B (*3*, *41*).

aDNA has also provided insights into the incidence and transmission of pathogens in Asia. Identification of Late Neolithic to Early Bronze Age *Yersinia pestis* strains, causative agent of the plague, in Bronze Age genetic outliers from the Cis-Baikal region is the first evidence for this lineage infecting individuals lacking Yamnaya-related steppe ancestry (*45*). Further reports of *Y. pestis* in Bronze Age sites in Yakutia and Cis-Baikal regions, representing the eastern-most Eurasian finding of this bacterium, may be indicative of broader spread of this disease in northeast Asia during the Bronze Age (*46*). aDNA studies have also highlighted the process of local adaptation to the extreme environmental conditions in the Tibetan Plateau and the Himalayas. Through the analysis of ancient remains, valuable insights into the timing of selection for high-altitude adaptation has provided a deeper understanding of how ancient populations successfully adapted to the challenging conditions of these regions (*9*, *10*).

Overall, ongoing exploration of aDNA holds great potential for uncovering further details about ancestral health and genetic influences on our current health landscape.

Community engagement in aDNA

With rapid advances in paleogenomics research, engaging with Indigenous and Descendant communities is crucial to advancing ethical, transparent, and meaningful studies. Community engagement is critical in Africa, given the predominant health and medical focus of

Box 2. Contributions of aDNA to the study of early population structure and movements in Asia and Oceania.

Upper Paleolithic human genomes, primarily from East Asia and Siberia, have augmented our understanding of the early diversification of modern human lineages out of Africa, their contributions to present-day genetic diversity, and introgression with Neanderthals and Denisovans (*56*). In the Holocene, the focus of most aDNA studies shifts toward broad-scale migrations across Asia and Oceania that are associated with subsistence and language expansions (*1*). For example, West Asia and the Near East have been central to questions about the origins of early farmers and their subsequent expansions (*1*). A parallel exists in East Asia, another major domestication center, where paleogenomics tracks decreasing genetic differentiation between northern and southern East Asia after the Neolithic, reflecting gene flow between these regions, as well as genetic connections between coastal Neolithic groups from Siberia through Japan to Southeast Asia (*56*). Similarly, several paleogenomic studies associated the movement of groups with language expansions during the Neolithic and Bronze Age in Eurasia. These include the origins and spread of Indo-European and Anatolian languages and Austronesian- and Austroasiatic-speaking groups migrating from East Asia to Island Southeast Asia and Oceania (*1*). Moreover, paleogenomics has traced long-range movements from East Asia to various regions of Oceania and the broader Pacific, some entailing multiple, diverse migration streams (*1*). Finally, aDNA has been used to investigate movements among Asia, Oceania, and the Americas. In northern East Asia, especially Siberia, studies have identified genetic lineages, such as the "Ancient North Siberians" and "Ancient Palaeo-Siberians," that are related to Native American founders, as well as later gene flow events between Siberia and North America (*57*).

population genetic studies in the region (*47*). As a result, population history–focused research on modern-day groups and aDNA studies are mainly conducted by nonlocal research institutions collaborating with local African institutions. Feedback to communities varies depending on previous engagement between them and collaborating institutions, with different teams adopting different approaches. Many aDNA projects in Africa were initiated through long-standing community-based archaeology projects spanning several decades [e.g., (*23*)] or established collaborations initially focused on modern-day population genetics research (*48*). Ethical concerns particularly relating to African aDNA research, the dignified treatment of human remains, and the protection of cultural heritage are important discussion topics and have been addressed in specific publications (*49*).

Paleogenomicists engaging with Indigenous communities in the Americas have produced substantial research focused on regional inquiries. One notable study resulted in the repatriation and reburial of ancestral remains (*11*), and another involved genomic analysis of contemporary tribal members and ancient individuals after several years of community engagement and dialogue with the Muwekma Ohlone Tribe from the San Francisco Bay (*13*). These mutually beneficial collaborative efforts between researchers and Indigenous communities yield valuable insights into local demographic events at a finer scale (*50*). Similarly, studies such as (*16*) that used aDNA techniques to facilitate repatriation of Ancestral remains with limited cultural contexts to Indigenous communities in Australia provide the impetus for similar work in other regions with a colonial past where families and Indigenous communities await the return of their Ancestors.

Box 3. Contributions of aDNA to the study of initial peopling and early dispersal of humans into the Americas.

The study of ancient genomes has contributed substantially to our understanding of how the Americas were initially peopled and how the first peoples dispersed throughout [reviewed in (*57*)]. The initial dispersal of peoples from northeast Asia into Beringia ~20,000 years ago was followed by an isolation period lasting a few millenia known as the Beringian Standstill. Humans then migrated south, likely through a coastal route ~17,000 years ago, followed by a rapid dispersal and branching into different populations across the continent. The first major branching separating Northern North Americans (NNAs) and Southern North Americans (SNAs) occurred ~15,000 to 16,000 years ago, with ubiquitous SNA ancestry detected in ancient and present-day populations from present-day US to Patagonia (*57*). South America was rapidly peopled by early SNA populations, followed by a continent-wide replacement by later SNA lineages at least 9000 years ago (*57*). The rapid dispersal of populations after the NNA-SNA split at different latitudes with contrasting environments implies that numerous adaptive and microevolutionary processes had to take place, resulting in local differentiation of human populations in the Holocene (*1*).

Challenges

With increasing focus on regional questions that tend to be inherently and closely intertwined with histories of present-day populations, especially Indigenous and Descendant communities, there is a need for tighter integration of genetics, archaeology, and community partnerships. Until this is achieved satisfactorily, current paleogenomic practices risk perpetuating imbalanced narratives without representation from all stakeholders, as well as widening the gap in scientific capacity between investigators in the Global North and the Global South.

Studies with a regional focus involving local researchers tend to more accurately integrate the voices of the local communities [e.g., (*11–13*)]. This is particularly important when Descendant communities belong to an Indigenous or marginalized community. There is a growing body of work exposing the discontent of Indigenous peoples, especially from the Global North, for being underrepresented and not considered in paleogenomic studies in which “samples” from their Ancestors are analyzed (*50*). As recognized by Bader *et al.*, this lack of engagement “emerges from the legacy of colonizing research in each of the fields [i.e. archaeology, biological anthropology, and genomics] from which paleogenomics draws” (*51*). Ultimately, the priorities and agendas of researchers, institutions, and grant and publication structures should not supersede those of communities. Meaningful engagement requires dedicated time and funds to develop and sustain partnerships with local communities (*51*). Nuances in community engagement exist, particularly in colonized contexts where oppression, murder, and forced relocations may make it less straightforward to identify Descendant communities with whom to engage.

Although the field of paleogenomics has grown by leaps and bounds since the publication of the first ancient human and Neanderthal genomes, most of the success of this field has been limited, scientifically and in terms of financial and global recognition, to groups from the Global North who lead these studies. This creates a massive imbalance in who designs and executes paleogenomics research, who decides priority questions in the field, and who ultimately benefits from the research. Building local capacity, both within scientific collaborations and Indigenous and Descendant community partnerships, has been suggested by both Indigenous researchers in the US (*51*) and researchers working in non-US contexts (*52*). It is generally acknowledged that efforts to be more inclusive, equitable, and engaged with local partners involve going beyond offering authorships. Researchers should consider avenues such as training and cowriting grants with local scientific collaborators, engaging regularly, asking for input on research questions, and returning results to Indigenous communities.

As local aDNA research expands, the need for larger sample sizes presents a substantial challenge within the field. It is crucial to emphasize the respect and preservation of human remains during the generation of aDNA data. Given the inherently destructive nature of generating aDNA data, it is important to consider the sequencing procedure for the sampled material, which is consumed during analysis. The two primary methods for extracting genetic information from ancient remains are capture sequencing and whole-genome shotgun sequencing. The widely used 1240K capture panel, initially introduced in (*53*) and accounting for ~70% of published ancient genomewide data as of 2023, has limitations in capturing rare variations in regions of low genetic diversity. Less biased capture panels that incorporate outgroup ascertainment schemes are now commercially available (*54*). However, the potential for limited overlap between these panels, along with the persistence of residual bias and batch effects, may hinder future comparative studies (*55*).

Although capture sequencing has long been considered a cost-effective option, the declining costs of whole-genome sequencing and the expenses associated with commercial capture panels have substantially narrowed the cost advantage. Low-coverage whole-genome sequencing of reasonably preserved human remains is comparable in price to commercial capture panels. Nevertheless, the key advantage of whole-genome sequencing remains its ability to provide comprehensive and unbiased genetic information, unlike the inherent bias introduced by capture panels. Whole-genome sequencing allows for the analysis of an individual’s complete genetic makeup, enabling direct population genetic investigations such as estimating population split times, assessing genetic diversity, and tracking changes in effective population size over time.

It is also worth noting that capture sequencing does not amplify metagenomes or enable the identification of human-associated microorganisms, including potential pathogens. To ensure the inclusion of metagenomes in analyses, it is important to encourage research groups to publish their complete raw metagenomic datasets alongside human genome–aligned binary alignment map (BAM) files. This call for standardized publication practices of

metagenomic data aims to enhance the comprehensiveness of aDNA studies and facilitate thorough evaluation by the scientific community.

Future directions and conclusion

Paleogenomics research, although a relatively new field, continues to captivate the attention of scientists and the general public because of its groundbreaking contributions to our understanding of the extensive population history of our species, spanning both ancient and more recent times. Although many studies have primarily focused on Europe and have addressed broader questions such as the out-of-Africa migration and the peopling of the Americas, this Review shifts its attention to regional inquiries outside of Europe that have yielded substantial knowledge and impact.

These highlighted studies conducted in the Americas, Asia, Oceania, and Africa share common themes that have greatly benefited from the use of aDNA. These themes include fine-scale demography and social structure, shedding light on aspects such as short-range migrations, regional genetic continuity, the influence of subsistence strategies on genetic diversification and time since divergence, as well as genetic admixture at local and community levels. Additionally, investigations into kinship and family structures through the study of pedigrees have contributed to our understanding of human societies in the past. Furthermore, the analysis of pathogen aDNA has exhibited excellent potential in revealing population movements and social interactions on a regional scale.

With recent advancements in wet lab techniques, DNA sequencing, and computational methods, the body of literature on aDNA is poised to expand substantially in the coming years. The exploration of both large-scale and regional questions, including those outlined in this Review, promises exciting frontiers that will further enhance our knowledge of deep time and recent human evolution. It is crucial that the field progresses with a strong foundation built upon ethical, engaged, and responsible research practices. These principles will serve as a cornerstone for the advancement of paleogenomics.

REFERENCES AND NOTES

1. M. Stoneking *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2209475119 (2023).
2. Y. Liu, X. Mao, J. Krause, Q. Fu, *Science* **373**, 1479–1484 (2021).
3. M. A. Spyrou, K. I. Bos, A. Herbig, J. Krause, *Nat. Rev. Genet.* **20**, 323–340 (2019).
4. F. Racimo, M. Sikora, M. Vander Linden, H. Schroeder, C. Lalueza-Fox, *Nat. Rev. Genet.* **21**, 355–366 (2020).
5. M. van de Loosdrecht *et al.*, *Science* **360**, 548–552 (2018).
6. R. Fregel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6774–6779 (2018).
7. M. Lipson *et al.*, *Nature* **577**, 665–670 (2020).
8. P. Gelabert *et al.*, *Curr. Biol.* **32**, 3232–3244.e6 (2022).
9. H. Wang *et al.*, *Sci. Adv.* **9**, eadd5582 (2023).
10. C.-C. Liu *et al.*, *Nat. Commun.* **13**, 1203 (2022).
11. J. Lindo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4093–4098 (2017).
12. A. Aqil *et al.*, *iScience* **26**, 106581 (2023).
13. A. L. Severson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2111533119 (2022).
14. M. Rasmussen *et al.*, *Nature* **506**, 225–229 (2014).
15. M. Rasmussen *et al.*, *Nature* **523**, 455–458 (2015).
16. J. L. Wright *et al.*, *Sci. Adv.* **4**, eaau5064 (2018).
17. V. Villa-Islas *et al.*, *Science* **380**, eadd6142 (2023).
18. M. R. Capodiferro *et al.*, *Cell* **184**, 1706–1723.e24 (2021).
19. C. de la Fuente *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4006–E4012 (2018).
20. N. Nakatsuka *et al.*, *Nat. Commun.* **11**, 3868 (2020).
21. J. Lee *et al.*, *Sci. Adv.* **9**, eadf3904 (2023).
22. R. Yaka *et al.*, *Curr. Biol.* **31**, 2455–2468.e18 (2021).
23. E. S. Brielle *et al.*, *Nature* **615**, 866–873 (2023).
24. S. Wasef *et al.*, *Front. Ecol. Evol.* **8**, 217 (2020).
25. N. E. Altınışık *et al.*, *Sci. Adv.* **8**, eabo3609 (2022).
26. I. Lazaridis *et al.*, *Science* **377**, 982–987 (2022).
27. C. Ning *et al.*, *Nat. Commun.* **11**, 2700 (2020).
28. C. Jeong *et al.*, *Cell* **183**, 890–904.e29 (2020).
29. F. Zhang *et al.*, *Nature* **599**, 256–261 (2021).
30. V. Shinde *et al.*, *Cell* **179**, 729–735.e10 (2019).
31. C. A. Fortes-Lima *et al.*, The genetic legacy of the expansion of Bantu-speaking peoples in Africa. bioRxiv 535432 [Preprint] (2023); https://doi.org/10.1101/2023.04.03.535432.
32. P. Skoglund *et al.*, *Cell* **171**, 59–71.e21 (2017).
33. M. E. Prendergast *et al.*, *Science* **365**, eaaw6275 (2019).
34. M. Steyn *et al.*, *Afr. Archaeol. Rev.* **40**, 53–72 (2023).
35. S. Marciniak, G. H. Perry, *Nat. Rev. Genet.* **18**, 659–674 (2017).
36. F. V. Seersholm *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7771–7776 (2018).
37. A. Grealy *et al.*, *J. Archaeol. Sci.* **75**, 82–88 (2016).
38. G. Kerner, J. Choin, L. Quintana-Murci, *Nat. Med.* **29**, 1048–1051 (2023).
39. S. Pfeiffer, L. Harrington, M. Lombard, *Int. J. Paleopathol.* **24**, 158–164 (2019).
40. R. F. Rifkin *et al.*, *Commun. Biol.* **6**, 240 (2023).
41. A. A. Guzmán-Solís *et al.*, *eLife* **10**, e68612 (2021).
42. T. P. Honap *et al.*, *Am. J. Biol. Anthropol.* ajpa.24735 (2023).
43. Å. J. Vågene *et al.*, *Nat. Commun.* **13**, 1195 (2022).
44. M. Bravo-Lopez *et al.*, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20190580 (2020).
45. H. Yu *et al.*, *Cell* **181**, 1232–1245.e20 (2020).
46. G. M. Kılınç *et al.*, *Sci. Adv.* **7**, eabc4587 (2021).
47. A. Choudhury, D. Sengupta, S. Aron, M. Ramsay, in *Africa, the Cradle of Human Diversity* (Brill, 2021), pp. 257–304; https://brill.com/view/book/9789004500228/BP000021.xml.
48. C. M. Schlebusch *et al.*, *Science* **358**, 652–655 (2017).
49. V. E. Gibbon, *Nat. Rev. Genet.* **21**, 645–647 (2020).
50. E. Kowal *et al.*, *HGG Adv.* **4**, 100161 (2023).
51. A. C. Bader *et al.*, in *Working With and for Ancestors*, C. H. Meloche, L. Spake, K. L. Nichols, Eds. (Routledge, 2020), pp. 166–177.
52. M. C. Ávila-Arcos, C. de la Fuente Castro, M. A. Nieves-Colón, M. Raghavan, *Front. Genet.* **13**, 880170 (2022).
53. I. Mathieson *et al.*, *Nature* **528**, 499–503 (2015).
54. P. Skoglund, "An 'ancestral 850k' SNP panel for studies of human genetic history using outgroup-ascertainment" (Arbor Biosciences, 2021); https://arborbiosci.com/wp-content/uploads/2021/03/Skoglund_Ancestral_850K_Panel_Design.pdf.
55. R. Davidson *et al.*, Allelic bias when performing in-solution enrichment of ancient human DNA. bioRxiv 547445 [Preprint] (2023); https://doi.org/10.1101/2023.07.04.547445.
56. M. Zhang, Q. Fu, *Curr. Opin. Genet. Dev.* **62**, 78–84 (2020).
57. E. Willerslev, D. J. Meltzer, *Nature* **594**, 356–364 (2021).

ACKNOWLEDGMENTS

The authors are profoundly thankful to M. Bravo-Lopez, E. Garfias-Morales, and E. Lomelin de Anda for their help with Fig. 1 and to E. Huerta-Sanchez for providing comments on the first version of the manuscript. **Funding:** M.C.A. is supported by the International Human Frontiers Science Program Organization (grant RGY0075/2019), the International Centre for Genetic Engineering and Biotechnology (grant CRP/MEX19-03), and the Universidad Nacional Autónoma de México Support Program for Research and Technological Innovation Projects (UNAM-PAPIIT) (grant IA203821). M.R. is supported by the National Institutes of Health (grant R35GM143094). C.S. is supported by the Knut and Alice Wallenberg Foundation.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 27 May 2023; accepted 5 September 2023
10.1126/science.adh8140

REVIEW

Ancient DNA genomics and the renaissance of herbaria

Hernán A. Burbano[1]*† and Rafal M. Gutaker[2]†

Herbaria are undergoing a renaissance as valuable sources of genomic data for exploring plant evolution, ecology, and diversity. Ancient DNA retrieved from herbarium specimens can provide unprecedented glimpses into past plant communities, their interactions with biotic and abiotic factors, and the genetic changes that have occurred over time. Here, we highlight recent advances in the field of herbarium genomics and discuss the challenges and opportunities of combining data from modern and time-stamped historical specimens. We also describe how integrating herbarium genomics data with other data types can yield substantial insights into the evolutionary and ecological processes that shape plant communities. Herbarium genomic analysis is a tool for understanding plant life and informing conservation efforts in the face of dire environmental challenges.

Herbarium specimens—meticulously collected and pressed plant samples—preserve a tangible record of botanical diversity and have long served as a foundation for botanical, taxonomical, and systematics studies (*1*). Hosted across 3000 herbaria, these resources encompass close to 390 million specimens and their associated metadata (*2*). We owe this number to centuries of specimen custodianship and current curation that ensures continuous growth, preservation, and sustainable use of this rich collection. Alongside their uses for botanical monographs and systematics (*3*), herbarium specimens covering diverse taxa and all continents increase the power of macroevolutionary studies, enabling investigations into trait evolution (*4*) and plant family radiations (*5*).

Recent advances in high-throughput sequencing technologies have facilitated the retrieval of genome-scale data from herbarium specimens (*6*). Herbarium-derived DNA can be classified as ancient DNA (aDNA) by its biochemical characteristics; it is highly fragmented, with distinct patterns of DNA misincorporations and breakage (*7*, *8*). Such patterns result from postmortem DNA damage, which is accelerated during specimen preparation (*9*). DNA damage does not reflect the evolutionary history of specimens, but it allows authentication of the historical nature of the herbarium-derived DNA (*7*, *10*). Additionally, herbarium-derived DNA originates not only from the plant host but also its associated commensal or pathogenic microbiomes. Distinguishing between the genuine microbiome and postmortem colonizers presents a methodological challenge (*11*), but, if resolved, herbarium metagenomics holds the potential to provide valuable insights into past biotic interactions.

Lack of cell and genome integrity makes functional and structural genomic tools such as chromatin conformation capture unsuitable for historical tissues, and the fragmented nature of DNA derived from herbarium specimens limits de novo assembly of full herbarium genomes with comparable quality to modern genomes. However, short-read sequencing approaches have been developed to generate genome-scale data from herbarium specimens (*12*, *13*). Although whole-genome sequencing allows genome-wide scans for selection to address questions related to the genetic basis of adaptation, it can be substituted by reduced-representation techniques for addressing population history questions. By contrast, cost-efficient techniques such as enrichment capture of conserved genes and plastome sequencing are informative for macroevolutionary and systematics studies. Amplicon sequencing approaches have been reviewed exhaustively elsewhere (*14*) and are outside the scope of this article.

> **"Herbarium collections offer valuable spatiotemporal data for investigating the genetic mechanisms underlying plant responses..."**

Despite the shortcomings associated with the degraded nature of herbarium-derived DNA, genomic studies of herbarium specimens offer a wide taxonomic and geographic sampling breadth to directly investigate the evolutionary responses of plants and their associated microbiomes over the past 300 years of anthropogenic global environmental change.

Contributions of herbarium genomics to taxonomic and evolutionary studies

Classification and macroevolutionary inference

The integration of sequencing data with morphological data has substantially increased our ability to group plants into taxonomic units. In botany, species definition comes from type specimens, which were used to describe the characteristics of species in the first place. Once the genetic composition of type specimens is established, a newly collected plant can be identified to the species level using molecular characterization alone. A widely used approach involves coupling genome skimming with chloroplast assembly or molecular barcode readout, allowing identification at the genus and species levels with high congruence with morphological identification (*15*). Such an approach was used to generate plastid and mitochondrial genomes from type specimens of red algae species (*Pyropia*) and allows molecular identification of unclassified material (*16*). Another challenge is to group species into higher taxonomic ranks, which often requires both morphological and genetic characterization. For example, genome-wide data helped resolve the challenging sweet potato genus (*Ipomea*) (*17*).

Herbarium collections harbor a wide range and diversity of species as well as geographies, which would not otherwise be accessible within the time frame of contemporary research projects. This is particularly useful for large-scale phylogenomic projects such as the Plant and Fungal Trees of Life, in which >28% of the taxa examined were sampled from herbarium specimens and were assessed using an enrichment capture approach (*18*). The resulting phylogenetic trees inform our understanding of macroevolution and the history of plant life on Earth. For example, this approach contributed to resolving the relationships of taxa in the mulberry family (*Moraceae*), and it indicated repeated, parallel evolution of inflexed stamens that facilitate wind pollination (*19*). Other uses include facilitating predictions of plant-based compounds, for example, the prediction of active compounds for medicinal use, a task in which herbarium collections play a special role (*20*).

Preserving a permanent record of the source materials from which DNA was retrieved is also important. Consequently, large biogenome projects, such as the Darwin Tree of Life, which aim to build reference genome assemblies for vast quantities of species, have partnered with institutions that generate voucher herbarium specimens for the plant taxa sequenced (*21*).

Tracing the evolutionary history of wild plants

Herbarium collections offer valuable spatiotemporal data for investigating the genetic mechanisms underlying plant responses to global environmental change, including the influence of modern agricultural practices (*22*, *23*). With

[1]Centre for Life's Origins and Evolution, Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK. [2]Royal Botanic Gardens, Kew, Kew Green, Richmond, Surrey TW9 3AE, UK.
*Corresponding author. Email: h.burbano@ucl.ac.uk
†These authors contributed equally to this work.

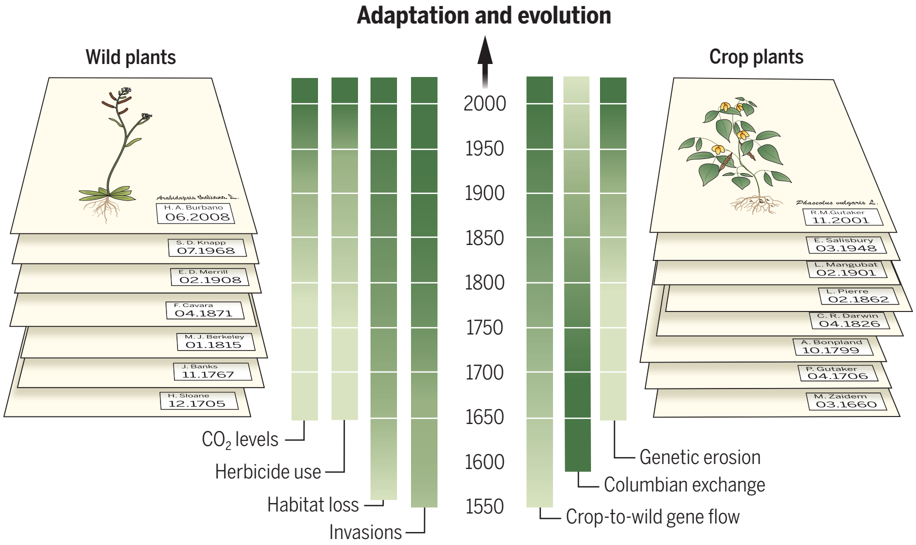


Fig. 1. The sampling range of herbarium specimens covers the time of intensified human influence on the environment. Sequencing the genomes of historical plants, as well as associated species, can address questions about their evolution and adaptation in response to anthropogenic pressures. Vertical bars indicate the intensity of human-mediated effects in the past five centuries (darker shades represent a stronger effect) on wild (left) and crop (right) plants. Stacks of herbarium specimens illustrate temporal sampling.

escalating global movement and trade, plants have become increasingly transported over long distances, heightening the likelihood of introduction of non-native species that might evolve into invasive species (*24*). Herbarium specimens have traditionally been used as proxies for documenting first-plant colonizers (Fig. 1). Sequences from herbarium specimens over dense time series can test hypotheses in invasion biology and determine whether changes in phenotypic traits are genetically encoded. Genetic analysis of herbarium specimens of garlic mustard (*Alliaria petiolata*) spanning 50 years of North American invasion suggested that differential secretion of phytochemicals across this time period had an adaptive genetic basis (*25*). Herbarium genomics also reconstructed the North American population history of a prevalent colonizing semiclonal lineage of thale cress (*Arabidopsis thaliana*) (*26*). A phylogenetic approach determined the time of colonization and connected the maintenance of de novo mutations in this lineage to root traits and flowering time that are likely important for the ecology of invasions. Human-induced alterations of landscapes has led to genetic mixing in the native range of common ragweed (*Ambrosia artemisiifolia*) in North America, before its introduction to Europe, potentially playing a role in its subsequent success as an invasive species (*27*). Furthermore, sequencing of numerous present-day and historical herbarium genomes of ragweed revealed that European populations exhibit a considerable turnover in genetic structure. This pattern is likely attributable to multiple introductions, founder effects, and genetic admixture with closely related European species (*28*).

Anthropogenic activities additionally generate habitats for wild plants. Agricultural fields face challenges from competition by weed species, requiring heavy use of herbicides to minimize interspecies competition and maintain crop yields. Excessive use of herbicides since their introduction in the 1950s has resulted in continuous evolution of weeds with resistance to these formerly deadly chemicals (Fig. 1) (*29*). Devising effective strategies for managing herbicide resistance necessitates determining whether resistance mutations originate de novo or were present in the existing genetic variation and to what extent gene flow plays a role in spreading herbicide resistance. Genotyping herbarium specimens of blackgrass (*Alopecurus myosuroides*) revealed that an herbicide resistance mutation located in the direct target of herbicides (target-site resistance) was segregating at a very low frequency approximately one century before the introduction of herbicides (*30*). A recent genomics study using contemporary samples also suggested that standing genetic variation plays a crucial role in facilitating the rapid evolution of target-site herbicide resistance (*31*). In addition to developing herbicide resistance, weeds must also adapt to the human-made environment of modern agricultural fields. A recent study of waterhemp (*Amaranthus tuberculatus*), an agricultural weed native to North America, used geographical pairing of hundreds of present-day genomes from natural and agricultural populations and a time series of hundreds of corresponding herbarium genomes (*32*). These data allowed monitoring of changes in allele frequencies over time as well as estimation of selection coefficients of variants observed predominantly in agricultural settings.

A time machine for recent crop evolution

As with wild plants, we can use time-stamped herbarium specimens of crop plants to address many important questions about their movement, adaptation, and changes in genetic diversity. Crop herbaria span a period of >500 years, with preliminary evidence showing that the oldest DNA sequences retrieved, from a tomato specimen (*Solanum lycopersicum*), date to ~1558 (*33*). This time interval overlaps with the Columbian "exchange" of crops globally, a serendipitous test for the limits of environmental adaptation (Fig. 1). This period saw en masse movement of European staples to colonies in the Americas, the introduction of American crops to Eurasia, and the global movement of cash crops for colonial exploitation. All continents were now connected by maritime routes, which resulted in rapid introduction of new crops, livestock, and diseases, with unprecedented effect on people's well-being (e.g., illness) and the environment (e.g., agricultural pressures) (*34*). Herbarium specimens helped settle a long debate about the origins and adaptation of the potato (*Solanum tuberosum*) when introduced from the Americas to Eurasia. Sequencing of herbarium specimens revealed the genetic turnover of introduced varieties over time (*35*) and showcased the expanding role of hybridization, with geographically distinct species that came into direct contact on the European continent (*36*). Putative preadaptation through phytohormonal regulation preceded an important change in day length–mediated tuberization that allowed potatoes to be cultivated in the completely different European photoperiod regime (*36*).

The past 250 years have seen increases in agricultural productivity. From the British Agricultural Revolution in the late 1700s to the Green Revolution in Latin America and South and Southeast Asia in the mid-1900s, new agrarian practices were accompanied by the adoption of new crop varieties (Fig. 1). Despite the general assumption that these processes had detrimental effects on the genetic diversity of crops, sequencing of seed bank materials has provided little evidence supporting massive genetic diversity loss (*37*). Genome sequencing of time-stamped crops from herbaria will yield a direct measure of genetic diversity change, together with other demographic parameters.

The postulated narrow gene pool available for crop breeding and the demand for genetic variants conferring stress tolerance, as well as recent developments in the field of plant

"neodomestication," have sparked interest in the genetic diversity of crop wild relatives (*38*). In addition to the detrimental effects of land use change and habitat fragmentation on genetic diversity, reports suggest that wild relatives of major crops are heavily influenced by gene flow from cultivated counterparts, even in self-pollinating species such as rice (*Oryza sativa*) (*39*). The possibility of gene flow raises concerns regarding genetically modified organisms, from which, for example, herbicide resistance variants could be passed to wild and weedy species. Comparative genomics of time-stamped herbarium specimens of crop wild relatives will provide a direct measure of crop-to-wild gene flow risks over centuries.

Herbarium specimens as time capsules of ecological interactions

Shotgun sequencing of herbarium specimens enables characterization of not just plant genomes but also genomes belonging to associated organisms. These include plant pathogens and the diverse microbial communities residing in the rhizosphere and phyllosphere, habitats for microorganisms on a plant's aerial surface and at the root–soil interface, respectively. Sporadically, other organisms such as insects, mollusks, and even small reptiles can be found in leaf creases. Therefore, herbarium specimens serve as windows into past biotic interactions, revealing coevolutionary trajectories between plants and their associated organisms (Fig. 2A).

This field initially focused on detecting filamentous crop pathogens, such as fungi and water molds (oomycetes), stemming from the availability of infected plant herbarium specimens. These were collected by early plant pathologists and are curated alongside micro and macro fungi in specialized collections known as fungaria. The Irish potato famine pathogen *Phytophthora infestans* was detected by polymerase chain reaction amplification of a short fragment of its mitochondrial DNA genome (*40*). Sequencing whole genomes from multiple historical samples considerably expanded the study of late blight, with a pandemic clonal lineage of *P. infestans* identified as the causative agent triggering the Irish potato famine in the mid-1800s (*41*, *42*). The strain from the 19th century was completely replaced by other strains in the 20th century, although the 19th-century mitochondrial DNA is still in circulation (*43*). Combined analysis of historical and modern samples ascertained that polyploidization distinguishes 20th-century samples from diploid 19th-century samples (*41*). Moreover, changes in allele frequencies of pathogen genes modulating infection that are recognized by the plant immune system—effector genes—were identified after the introduction of resistant potato cultivars in the 20th century (*41*).

Bacteria can also be accurately identified through analysis of herbarium-derived shotgun reads. A proof-of-principle study using a library preparation method that specifically enriches aDNA molecules successfully assembled de novo genomes for commensal and pathogenic bacteria belonging to the genus *Pseudomonas* from a 19th-century potato herbarium specimen (*44*). Herbarium-derived genomes of *Xanthomonas citri* pv. *citri* (*Xci*), the causal agent of Asiatic citrus canker, have been used to reconstruct the population history of this bacterial pathogen at both local (*45*) and global scales (*46*), revealing that the spread of Asian citrus canker disease is linked to climatic changes and the onset and global expansion of citriculture (*46*).

Whereas agricultural disease outbreaks are often characterized by expansion of individual pathogenic lineages, wild plants are infected by genetically diverse groups of pathogens (*47*). Preliminary data annotating bacteriophages in present-day and herbarium-derived *Pseudomonas* genomes identified a putative bacteriophage-derived bacteriocin (tailocin) used by pathogenic bacteria for intralineage competition, which likely prevents expansion of single bacterial lineages (*48*). This study showcases the resolution attainable through analysis of herbarium-derived bacterial genomes: determining the genomic location of the tailocins within *Pseudomonas* genomes, characterizing the

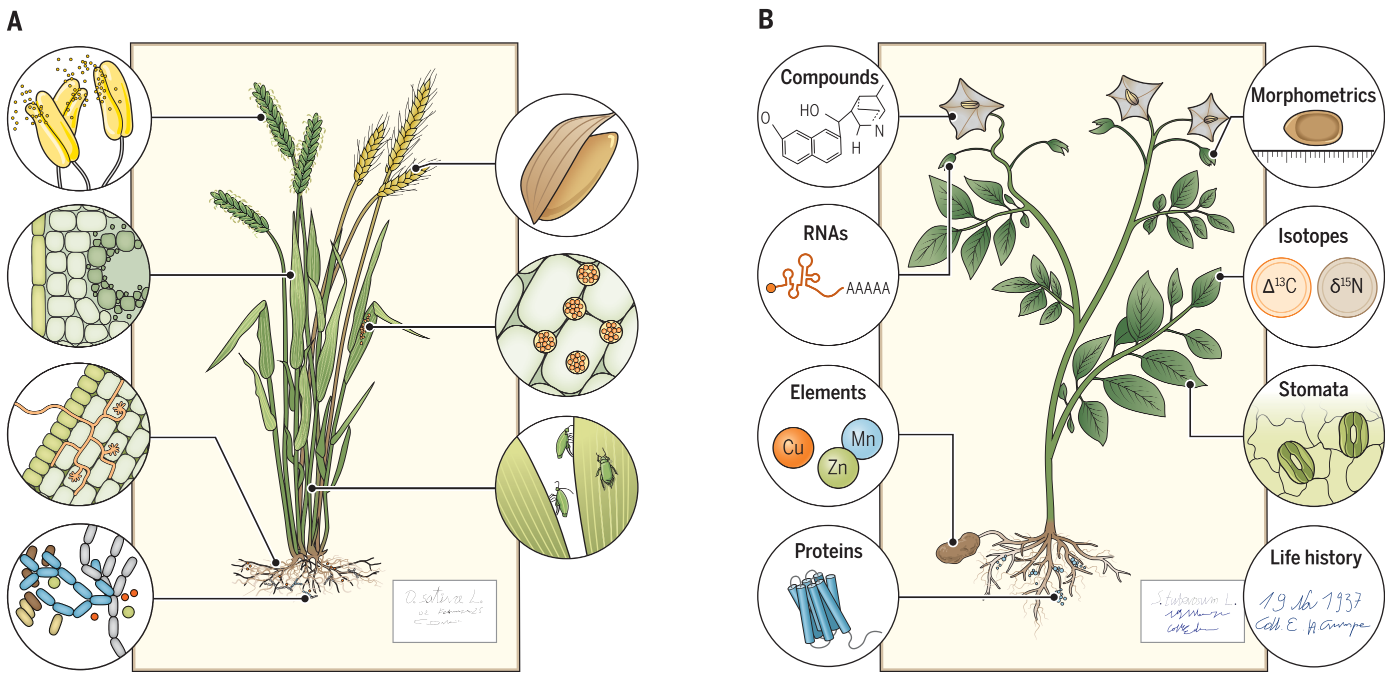


Fig. 2. Environmental, molecular, and phenotypic wealth of herbarium specimens. (**A**) Herbarium sheets preserve not only the genomes of individual plants but also those of many associated species: microbes, such as bacteria and fungi, as well as macroorganisms, such as insects. (**B**) Herbarium specimens also harbor a plethora of other types of data. Phenology and, occasionally, environmental or ethnobotanical information is recorded on the label. Anatomical features allow direct morphometric measurements. Other molecular data are available, from simple micronutrients and atomic isotopes to complex molecules such as alkaloids, RNAs, and proteins.

CREDIT: K. HOLOSKI/*SCIENCE* BASED ON RAFAL M. GUTAKER

tailocin gene content, assessing conserved colinearity of tailocin genes, and elucidating the haplotypes of highly diverse proteins likely crucial for tailocin bactericidal activity (*48*). Outside of pathogenic microbes, a targeted genomics approach was used to sequence and assemble symbiotic bacteria of wild yam (*Dioscorea sansibarensis*), revealing the horizontal transmissions of symbionts (*49*).

Studies of fungi, oomycetes, and bacteria have focused on analysis of pathogen-infected herbarium tissue or highly abundant microbial taxa expected to be found on plant tissues according to reference microbiomes retrieved from fresh plant samples. Pure metagenomic assignment of herbarium-derived sequencing reads is complicated by short read length and nucleotide misincorporation, which decrease the reliability of read mapping. Once reads are reliably assigned to taxa, subsequent challenges lie in distinguishing between the authentic microbiome and postmortem colonizers. Investigating whether patterns of DNA misincorporation in reads mapped to specific microbial genomes resemble those commonly observed in aDNA (*50*) or using library preparation methods that selectively incorporate aDNA reads (*44*) is insufficient to distinguish the authentic microbiome. It is crucial to assess the significance of identified taxa by comparing them to reference microbiomes from fresh plant tissues and herbarium contaminants. A promising alternative involves characterizing microbial taxa preferentially found in herbarium specimens across different plant taxa and multiple herbaria locations (*11*). This approach makes it possible to differentiate the authentic metagenome from potential contaminants within the herbarium metagenome. Shotgun metagenomics was used on present-day and herbarium genomes of *A. artemisiifolia* to characterize the pathogen load of plants inhabiting invasive and native ranges. A lower prevalence of disease-inducing plant pathogens in the invasive range suggests that escape from naturally occurring pathogens contributed to the invasion of this species in Europe (*28*).

Herbarium specimens retain plant roots, a habitat rich in microorganisms, and in certain instances contain remnants of the original soil (Fig. 2A). Characterizing root microbiomes could potentially unveil the dynamics of microbial communities influenced by land management or agricultural practices, such as the application of agrochemicals.

Beyond herbarium genomics: Integrating genomics with other data

Combining herbarium genomics with other molecular data

Herbaria have also proven valuable to scientists from other plant biology disciplines and for research on global environmental change (*22*, *23*). These alternative applications encompass the use of multiproxy molecular data. The biggest challenge in using molecules other than DNA from historic materials lies both in the limited understanding of their degradation and in the lack of authentication procedures. Although simple elements and their stable isotopes are not expected to degrade within herbarium timescales, care must be taken to avoid contamination. Ecologists as well as soil and plant scientists assess the state of ecosystems using nitrogen concentrations and isotope ratios, which reflect the cycling of key chemical elements. Herbarium specimens of multiple species allow monitoring of changes in nitrogen availability on a regional scale (Fig. 2B). Despite increased anthropogenic nitrogen deposition, foliar nitrogen concentrations have decreased (*51*). Combining analysis of nitrogen and carbon isotopes with phenology traits reveals heterogeneous physiological responses of *A. thaliana* that depend on growth habit, population structure, and changes in the environment (*52*). Combining such studies with genomic characterization might reveal the extent to which responses to anthropogenic pressures are genetic or plastic. This possibility is particularly relevant for crops, whose genetic makeup changed contemporaneously with the onset of fertilization regimes during the Green Revolution.

Complex chemical molecules produced by plants can modulate biotic interactions such as plant defense, plant-plant competition, and pollinator attraction. Thus, characterizing such molecules through time can reveal the evolution of multiple biotic interactions. However, isolating and quantifying complex chemical molecules is more challenging than isolating DNA and requires fine-tuning of multiple specialized approaches. It is important to understand the kinetics of each quantified molecule, including its half-life and degradation products. A study comparing terpenoids, important metabolites for growth and defense, in different sages (*Salvia* sp.) noted a significant effect of species but not of herbarium collection date on the terpenoid composition, although concentrations decrease over time. Therefore, herbarium specimens are suitable for qualitative comparison of terpenoids, but care needs to be taken for their quantitative analysis (*53*). Specimens of some plant species were annotated with their historical chemical composition at the time of collection, allowing direct comparison with modern measurements. For example, quinine alkaloid concentrations measured 150 years ago in fever tree (*Cinchona* sp.) barks are broadly congruent with recent measurements, suggesting that this group of compounds is stable (*54*). Degradation dynamics have been assessed for only a handful of biomolecules and should be evaluated individually for economically important plant products (medicinal, plant defense). Direct measurement coupled with genome sequencing should facilitate discovery of underlying genetic pathways (Fig. 2B).

Proteomes have been successfully isolated from ancient animal samples. Proteins degrade more slowly than DNA, and strict protocols for authenticating their historical or ancient nature have already been developed (*55*). However, these approaches have not yet been applied to herbarium specimens. Although RNA is much less stable than DNA, and authentication procedures are not currently available, its sequencing can be authenticated indirectly through damage pattern signatures in complementary DNA, providing, for example, insights into plant–microbe interactions at the time of sampling.

Leveraging phenotypic and genomic data from herbarium specimens

Examining changes in the timing of phenological milestones such as leaf-out or flowering allows the effects of climate change to be studied. Herbarium specimens are valuable and reliable records for phenological research (Fig. 2B). Automation of phenological scoring through machine learning approaches will expand the scope of herbarium-based phenological studies, enabling assessment of more phenotypes (*56*). The collection date of flowering herbarium specimens provides a reliable estimation of past flowering time (*57*). For example, long-term phenology trends from herbarium specimens indicate that European forest wildflowers shifted their phenology in response to climate change (*58*). However, any genetic basis is yet to be determined. Integrating herbarium genomics with phenological research will reveal whether variation in phenotypic traits has a genetic basis or is solely attributable to phenotypic plasticity.

Another important phenotype in the context of global environmental change is stomatal density, a measure of the pores that plants use for gas exchange (Fig. 2B). Stomatal density is expected to decrease as atmospheric CO_2 levels rise with industrialization, as has been revealed by herbarium specimens (*59*). A polygenic score–like metric that was devised by integrating herbarium and present-day genomes with known effects of gene knockout mutations on stomatal development preliminarily showed a positive correlation with stomatal density in *A. thaliana* (*60*). This metric was validated using present-day genomes and, when applied to herbarium genomes, reproduces previously observed patterns of stomatal density decrease over time. This correlation was achieved without directly measuring stomatal density in herbarium specimens, indicating a potential genetic basis underlying the observed patterns (*60*).

Characterization of interactions between plants and insect herbivores is enabled by quantifying lesions on herbarium specimens (Fig. 1). Analysis of herbarium specimens spanning

more than a century indicates that damage to plants by insect herbivores will likely continue to increase in the northeastern US as global temperatures rise (*61*). Herbarium genomics might enhance these studies by identifying herbivores through insect DNA left on plant tissue as well as plant genetic loci responsible for susceptibility to insect herbivory.

Outlook

Herbarium genomic approaches reveal the profound influence of human activities such as agriculture on plant evolution. However, there is still much to explore, particularly regarding anthropogenic effects on plant interactions with the environment. Biotic interactions can be investigated through metagenomics of herbarium specimens, whereas isotope ratios, stomata numbers, and phenology provide insights into abiotic interactions. Although the influence of climate change on population diversity and structure is routinely established, its evolutionary consequences and adaptive responses are rarely investigated (*62*). Integration of present-day and herbarium genomics continues to be essential for studying rapid adaptation in natural populations.

With time, new ways to use herbarium collections emerge, and their potential grows with the introduction of new analytical techniques. For example, recently, herbarium specimens have gained relevance for plant "de-extinction," the resurrection of plants considered extinct in the wild using preserved seeds or tissues from herbarium collections (*63*). Herbarium genomics holds potential for informing and validating such resurrection efforts.

Finally, herbarium collections temporally overlap colonial history, reflecting the acquisition and preservation of botanical treasures from continents explored by Europeans since the 15th century (*64*). Most specimens were obtained without proper permits or consent from Indigenous communities. Today, legal and ethical obligations regarding herbarium materials are guided by the Convention on Biological Diversity (CBD) and the benefit-sharing principles of the Nagoya Protocol. Although the CBD applies primarily to recent collections, ethical considerations extend to older materials, particularly plants cultivated on Indigenous lands, including crops, medicinal plants, and sacred species.

REFERENCES AND NOTES

1. V. A. Funk, "100 uses for an herbarium (well at least 72)," *American Society of Plant Taxonomists Newsletter*, vol. 17, issue 2, pp. 17–19 (2003); https://repository.si.edu/bitstream/handle/10088/11385/bot_2003_pr_Funk_100Uses.pdf.
2. B. M. Thiers, Index Herbariorum [continuously updated] (Steere Herbarium, New York Botanical Garden, 2022); https://sweetgum.nybg.org/science/ih/.
3. O. M. Grace *et al*., *Trends Plant Sci.* **26**, 433–441 (2021).
4. J. M. Heberling, *Int. J. Plant Sci.* **183**, 87–118 (2022).
5. S. Ramírez-Barahona, H. Sauquet, S. Magallón, *Nat. Ecol. Evol.* **4**, 1232–1238 (2020).
6. L. Kistler *et al*., *Annu. Rev. Plant Biol.* **71**, 605–629 (2020).
7. C. L. Weiß *et al*., *R. Soc. Open Sci.* **3**, 160239 (2016).
8. B. Shapiro, M. Hofreiter, *Science* **343**, 1236573 (2014).
9. M. Staats *et al*., *PLOS ONE* **6**, e28448 (2011).
10. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, *Bioinformatics* **29**, 1682–1684 (2013).
11. V. C. Bieker *et al*., *Mol. Ecol. Resour.* **20**, 1206–1219 (2020).
12. S. Dodsworth *et al*., *Trends Plant Sci.* **24**, 887–891 (2019).
13. S. M. Latorre, P. L. M. Lang, H. A. Burbano, R. M. Gutaker, *Curr. Protoc. Plant Biol.* **5**, e20121 (2020).
14. V. C. Bieker, M. D. Martin, *Bot. Lett.* **165**, 409–418 (2018).
15. P. G. Nevill *et al*., *Plant Methods* **16**, 1 (2020).
16. J. R. Hughey *et al*., *Sci. Rep.* **4**, 5113 (2014).
17. P. Muñoz-Rodríguez *et al*., *Nat. Plants* **5**, 1136–1144 (2019).
18. W. J. Baker *et al*., *Syst. Biol.* **71**, 301–319 (2022).
19. E. M. Gardner *et al*., *Taxon* **70**, 946–988 (2021).
20. O. A. Pérez-Escobar *et al*., *Science* **369**, 781–782 (2020).
21. M. K. N. Lawniczak *et al*., *Wellcome Open Res.* **7**, 187 (2022).
22. E. K. Meineke, C. C. Davis, T. J. Davies, *Ecol. Monogr.* **88**, 505–525 (2018).
23. P. L. M. Lang, F. M. Willems, J. F. Scheepens, H. A. Burbano, O. Bossdorf, *New Phytol.* **221**, 110–122 (2019).
24. M. van Kleunen *et al*., *Nature* **525**, 100–103 (2015).
25. R. A. Lankau, V. Nuzzo, G. Spyreas, A. S. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 15362–15367 (2009).
26. M. Exposito-Alonso *et al*., *PLOS Genet.* **14**, e1007155 (2018).
27. M. D. Martin *et al*., *Mol. Ecol.* **23**, 1701–1716 (2014).
28. V. C. Bieker *et al*., *Sci. Adv.* **8**, eabo5115 (2022).
29. J. M. Kreiner, J. R. Stinchcombe, S. I. Wright, *Annu. Rev. Plant Biol.* **69**, 611–635 (2018).
30. C. Délye, C. Deulvot, B. Chauvel, *PLOS ONE* **8**, e75117 (2013).
31. S. Kersten *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2206808120 (2023).
32. J. M. Kreiner *et al*., *Science* **378**, 1079–1085 (2022).
33. R. Vos *et al*., bioRxiv 2022.06.07.495125 [Preprint] (2022); https://doi.org/10.1101/2022.06.07.495125.
34. A. W. Crosby, *The Columbian Exchange: Biological and Cultural Consequences of 1492* (Greenwood Publishing Group, 2003).
35. M. Ames, D. M. Spooner, *Am. J. Bot.* **95**, 252–257 (2008).
36. R. M. Gutaker *et al*., *Nat. Ecol. Evol.* **3**, 1093–1101 (2019).
37. M. van de Wouw, T. van Hintum, C. Kik, R. van Treuren, B. Visser, *Theor. Appl. Genet.* **120**, 1241–1252 (2010).
38. A. Alsantely *et al*., *J. Exp. Bot.* **74**, 1331–1342 (2023).
39. H. Wang, F. G. Vieira, J. E. Crawford, C. Chu, R. Nielsen, *Genome Res.* **27**, 1029–1038 (2017).
40. J. B. Ristaino, C. T. Groves, G. R. Parra, *Nature* **411**, 695–697 (2001).
41. K. Yoshida *et al*., *eLife* **2**, e00731 (2013).
42. M. D. Martin *et al*., *Nat. Commun.* **4**, 2172 (2013).
43. M. D. Martin, S. Y. W. Ho, N. Wales, J. B. Ristaino, M. T. P. Gilbert, *Mol. Biol. Evol.* **31**, 1414–1420 (2014).
44. C. L. Weiß, M.-T. Gansauge, A. Aximu-Petri, M. Meyer, H. A. Burbano, *BMC Genomics* **21**, 432 (2020).
45. P. E. Campos *et al*., *PLOS Pathog.* **17**, e1009714 (2021).
46. P. E. Campos *et al*., *Nat. Commun.* **14**, 4306 (2023).
47. T. L. Karasov *et al*., *Cell Host Microbe* **24**, 168–179.e4 (2018).
48. T. Backman *et al*., bioRxiv 2023.04.17.536465 [Preprint] (2023); https://doi.org/10.1101/2023.04.17.536465.
49. B. Danneels *et al*., *Curr. Biol.* **31**, 2666–2673.e4 (2021).
50. F. M. Key, C. Posth, J. Krause, A. Herbig, K. I. Bos, *Trends Genet.* **33**, 508–520 (2017).
51. K. K. McLauchlan, C. J. Ferguson, I. E. Wilson, T. W. Ocheltree, J. M. Craine, *New Phytol.* **187**, 1135–1145 (2010).
52. V. L. DeLeo, D. N. L. Menge, E. M. Hanks, T. E. Juenger, J. R. Lasky, *Glob. Change Biol.* **26**, 523–538 (2020).
53. I. J. Foutami, T. Mariager, R. Rinnan, C. J. Barnes, N. Rønsted, *Front. Plant Sci.* **9**, 1877 (2018).
54. N. A. Canales *et al*., *J. Ethnopharmacol.* **249**, 112375 (2020).
55. J. Hendy *et al*., *Nat. Ecol. Evol.* **2**, 791–799 (2018).
56. K. D. Pearson *et al*., *Bioscience* **70**, 610–620 (2020).
57. C. C. Davis, C. G. Willis, B. Connolly, C. Kelly, A. M. Ellison, *Am. J. Bot.* **102**, 1599–1609 (2015).
58. F. M. Willems, J. F. Scheepens, O. Bossdorf, *New Phytol.* **235**, 52–65 (2022).
59. F. I. Woodward, *Nature* **327**, 617–618 (1987).
60. P. L. M. Lang *et al*., bioRxiv 2022.10.23.513440 [Preprint] (2022); https://doi.org/10.1101/2022.10.23.513440.
61. E. K. Meineke, A. T. Classen, N. J. Sanders, T. Jonathan Davies, *J. Ecol.* **107**, 105–117 (2019).
62. R. D. Clark *et al*., *Mol. Ecol. Resour.*1755-0998.13789 (2023).
63. G. Albani Rocchetti *et al*., *Nat. Plants* **8**, 1385–1393 (2022).
64. D. S. Park *et al*., *Nat. Hum. Behav.* **7**, 1059–1068 (2023).

ACKNOWLEDGMENTS

We thank P. Kersey and S. Latorre for their critical appraisal of the manuscript. **Funding:** H.A.B. was supported by the Royal Society (RSWF\R1\191011) and the Leverhulme Trust (Philip Leverhulme Prize). R.M.G. was supported by UKRI Frontier Research Guarantee (EP/X022404/1). **Competing interests:** None declared. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 1 June 2023; accepted 2 August 2023
10.1126/science.adi1180

REVIEW

Deep-time paleogenomics and the limits of DNA survival

Love Dalén[1,2,3]†, Peter D. Heintzman[1,4]†, Joshua D. Kapp[5,6]†, Beth Shapiro[6,7]*†

Although most ancient DNA studies have focused on the last 50,000 years, paleogenomic approaches can now reach into the early Pleistocene, an epoch of repeated environmental changes that shaped present-day biodiversity. Emerging deep-time genomic transects, including from DNA preserved in sediments, will enable inference of adaptive evolution, discovery of unrecognized species, and exploration of how glaciations, volcanism, and paleomagnetic reversals shaped demography and community composition. In this Review, we explore the state-of-the-art in paleogenomics and discuss key challenges, including technical limitations, evolutionary divergence and associated biases, and the need for more precise dating of remains and sediments. We conclude that with improvements in laboratory and computational methods, the emerging field of deep-time paleogenomics will expand the range of questions addressable using ancient DNA.

The Pleistocene epoch [approximately 2.6 million years ago (Ma) to 10 thousand years ago (ka)] was a time of considerable environmental upheaval that shaped the present worldwide distribution of biodiversity. Environmental changes during the Pleistocene included cyclical fluctuations in global temperatures and precipitation patterns, advances and recessions of high-latitude ice sheets, and substantial changes in sea level, together with large-scale volcanism, paleomagnetic reversals, and the global spread of humans (*1*). These events altered habitats around the world, driving changes in resource availability and ecological community composition.

The rich fossil record of the Pleistocene has been instrumental for testing hypotheses about correlation between these environmental changes and biodiversity dynamics, especially at high latitudes where the cold climate favors fossil preservation. This is particularly true for the Late Pleistocene (126 to 11.7 ka), thanks to fine-scale inferences enabled by ancient DNA preserved in fossils dating to this period. Such inferences have allowed insights into population turnover (*2–4*) and interspecies gene flow (*5*)—processes that are invisible to traditional paleontological techniques—and shown that demographic trends in large mammals closely track available habitat (*6*).

Technical advances in DNA recovery have extended the ability to make these inferences deeper into the Pleistocene. DNA from bones and teeth that are several hundreds of thousands of years old (*7–9*) and beyond one million years old (*10*) has now been recovered and analyzed (Fig. 1). Such deep-time paleogenomes, which we consider here to refer to genomes assembled from organisms that lived during or earlier than the Middle Pleistocene, i.e., >126 ka, are still rare because postmortem processes lead to successive degradation of DNA molecules into increasingly small fragments,

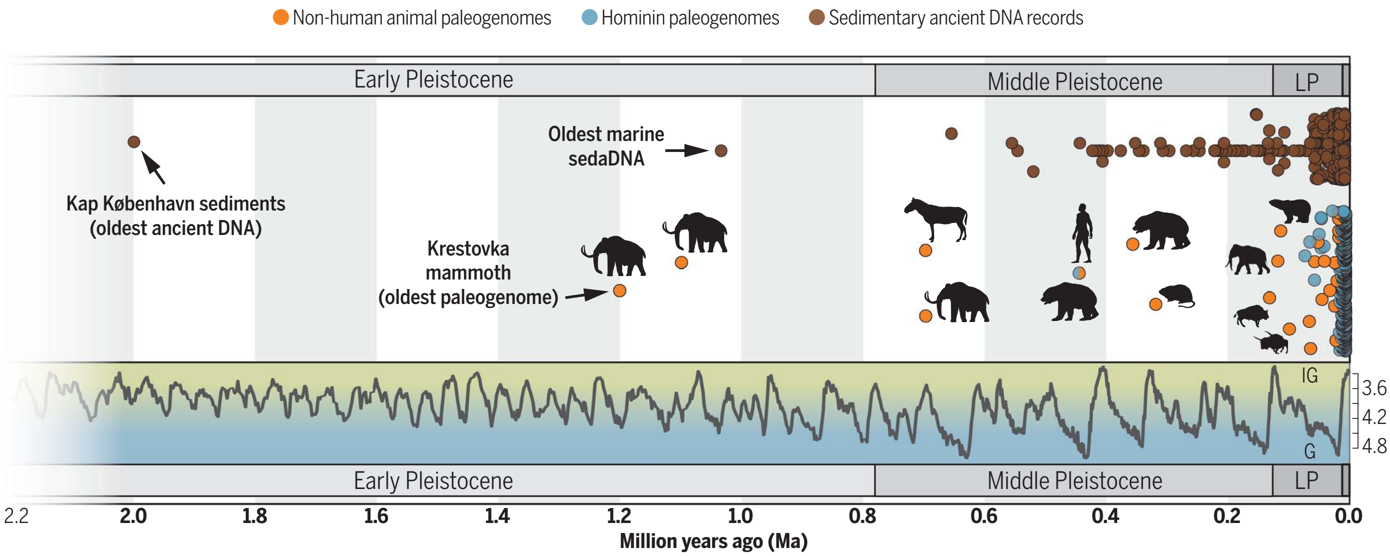


Fig. 1. The temporal distribution of ancient DNA studies to date highlights gaps and opportunities for deep-time paleogenomics and sedimentary ancient DNA. Most ancient DNA studies fall within the last 50 ka and the most recent glacial cycle. The climate curve is based on benthic δ^{18}-oxygen measurements [per mil; LR04 stack from (*42*)]. Sedimentary ancient DNA data are from the AncientMetagenomeDir (v23.06.0) (*57*) and (*58*), with metabarcoding records older than one million years excluded. Paleogenomic data are available from (*59*). Paleogenomes older than 100 ka are annotated with a silhouette of the study taxon, with the deep-time paleogenomes including a 130-ka steppe bison (*36*); 330-ka collared lemming (*40*); 360-ka cave bear (*9*); 430-ka cave bear and hominin (*35*, *60*); 700-ka horse (*8*); and 700-ka, 1.1-Ma, and 1.2-Ma mammoths (*10*). LP, Late Pleistocene; IG, Interglacial; G, Glacial.

[1]Centre for Palaeogenetics, Svante Arrhenius väg 20C, SE-10691 Stockholm, Sweden. [2]Department of Zoology, Stockholm University, SE-10691 Stockholm, Sweden. [3]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, SE-10405 Stockholm, Sweden. [4]Department of Geological Sciences, Stockholm University, SE-10691 Stockholm, Sweden. [5]Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA, 95064, USA. [6]Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA, 95064, USA. [7]Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, CA, 95064, USA.
***Corresponding author. Email: bashapir@ucsc.edu**
†These authors contributed equally to this work.

making DNA recovery more difficult with age. Early and Middle Pleistocene DNA has, however, been recovered from remains and sediments in high-latitude permafrost (*10–14*) and lower latitude caves (*15*, *16*), suggesting that deep-time genomics is feasible in ideal preservation environments. Here, we explore the current state of the art in deep-time paleogenomics research, the key obstacles preventing wider adoption, and scientific questions that deep-time paleogenomics can address.

DNA persistence into deep time

DNA does not survive indefinitely but it does survive for considerably longer than predicted by the earliest models. In 1993, Lindahl estimated that hydrolytic depurination would lead to complete degradation of DNA molecules within several tens of thousands of years (*17*). This limit has since been exceeded, and DNA is regularly recovered from remains and sediments that date to within the last 100 ka. As of September 2023, the oldest reconstructed paleogenome is from a permafrost-preserved mammoth dating to between 1 and 2 Ma (*10*) and the oldest isolated DNA is from ~2-Ma sediment from northern Greenland (*11*). However, the maximum age of recoverable and useful DNA molecules—those that are long enough to retain information—remains uncertain.

DNA begins to degrade immediately following organismal death, initially through microbial and endogenous nuclease activity (Fig. 2). In nuclear DNA, strands are cleaved in labile regions of histone-DNA complexes, resulting in a ~10-base periodicity in the distribution of the lengths of recovered molecules (*18*). The primary chemical mechanism of DNA fragmentation is hydrolytic depurination. This process removes adenine or guanine bases, creating abasic sites that can be cleaved by β elimination (*19*) (Fig. 2C), and leading to purine overrepresentation adjacent to strand breaks (*20*) (Fig. 2E) and interior gaps (*21*). Hydrolytic deamination, another common form of chemical damage, converts cytosine to uracil and is observed as thymine in sequencing data, or "C-to-T transitions" (Fig. 2C). Deamination occurs primarily near strand ends and in single-stranded DNA (*17*, *21*, *22*) (Fig. 2E). DNA crosslinking (*19*, *22*) and oxidative damage (*20*, *23*) also occur but are observed less frequently than depurination and deamination. These typical damage patterns can be used to bioinformatically corroborate the authenticity of recovered ancient sequences and, to reduce their impact on sequence accuracy, can be identified and removed from ancient DNA data sets using standard bioinformatic approaches.

Recovery of increasingly old and damaged DNA is possible in part due to technical advances in the laboratory. Ancient DNA isolation methods are optimized to recover both short DNA molecules and molecules containing nicks and

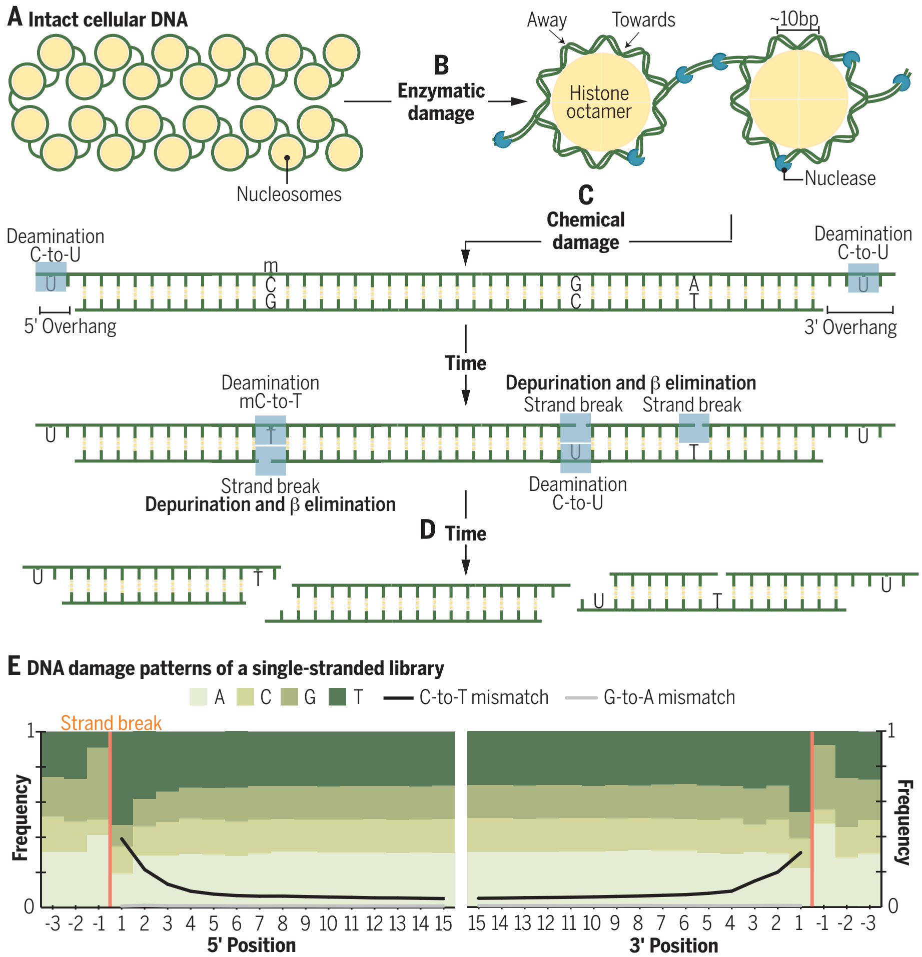


Fig. 2. DNA fragmentation and degradation begins after death and continues until fragments are too short to be useful. (**A**) The integrity of megabase length DNA is maintained by a cell's enzymatic repair machinery and, in eukaryotic genomes, packaged in histone-DNA complexes. (**B**) Following death, repair stops and DNA damage begins to accumulate. Nucleases and microorganisms cleave DNA in labile regions between nucleosomes and when the DNA backbone faces away from histones. (**C**) Over time, chemical damage also accumulates. Cytosine bases are converted to uracil and methylated cytosines are converted to thymines (by deamination). Cytosines are particularly vulnerable to deamination in single-stranded regions such as in overhanging regions at DNA termini, but deamination is possible in some double-stranded contexts. Fragmentation occurs after the loss of purine bases (depurination), creating abasic sites that can be cleaved by β elimination. Depurination and β elimination create a region of single-stranded DNA, which leaves cytosines vulnerable to deamination. (**D**) Given enough time, DNA molecules will become too short to be identifiable. (**E**) A summary of base and mismatch frequencies along the initial 15 5′ and 3′ bases of reads generated using a single-stranded DNA library protocol [library JKFC14; (*25*)]. Depurination leads to overrepresentation of adenine and guanine bases adjacent to strand breaks. C-to-T mismatches are elevated near read ends and observed throughout damaged reads. Whereas 3′ G-to-A mismatches are observed in double-stranded libraries, single-stranded libraries show a C-to-T signal at both ends by retaining the native termini of the molecules.

gaps. Extracted molecules are prepared for sequencing by ligating platform-specific adapters to either double-stranded or single-stranded DNA. Single-stranded approaches to genomic library preparation (*24*, *25*) convert natively single-stranded DNA as well as double-stranded DNA and more effectively convert molecules containing nicks and gaps compared with double-stranded approaches. DNA extracts are also often treated with uracil DNA glycosylase and endonuclease VIII to reduce deamination damage by removing uracil bases (*26*). Although this approach reduces damage-induced errors in the resulting sequencing data, it also cuts the DNA backbone at abasic sites and shortens the recovered molecules by 5 to 10 nucleotides (*26*). Deep-time DNA molecules are already short—often <35 bases (*15*)—and consequently this may reduce the proportion of useful endogenous DNA.

The short nature of deep-time DNA molecules makes them prone to spurious alignment and reference bias (*27*), complicating genome assembly and analysis. For example, ancient DNA data sets comprise both endogenous DNA from the target organism(s) and introduced exogenous DNA. These categories of molecules can be separated by identifying each read through taxonomic assignment, which can be problematic if the ancient organism has no close living relative to act as a genomic reference. Lack of a close reference, reference bias, and errors introduced by damage will also impede variant and consensus calling. Bioinformatic approaches mitigate these challenges by directly modeling DNA damage and/or bias as part of genotyping (*28*) or considering only substitutions that are not affected by cytosine deamination. Reference genomes can also be modified to create artificially closer references, such as a "Neanderthalized" version of the human reference genome for reference-guided mapping of Neanderthals reads (*29*). Genotype likelihoods rather than strictly called genotypes can also be used during downstream analysis, although imputation-based analytical methods may be inappropriate for deep-time data sets if ancient genomic diversity is not represented in existing reference panels.

Research opportunities arising from deep-time DNA

Speciation and evolution

Speciation is not always a simple process of cladogenesis followed by reproductive isolation. Instead, modern and paleogenomic data have shown that interspecific hybridization is surprisingly common and perhaps driven in part by repeated habitat redistribution associated with glacial cycles (*5*, *9*, *10*). For example, brown bears and polar bears hybridized during previous glacial and interglacial periods (*30*, *31*) as well as in the modern era. Recently, polar bear and cave bear paleogenomes dating to up to 360 ka revealed that all living brown bears derive a portion of their ancestry from admixture with these other bear lineages—evolutionary events that were invisible without these paleogenomes (*9*, *32*). Similarly, a mammoth paleogenome dating to the Early Pleistocene revealed that Columbian mammoths (*Mammuthus columbi*) originated after hybridization between two distinct ancient mammoth lineages (*10*) (Fig. 3). Taxonomically diverse deep-time paleogenomes could clarify the timing, rate, and extent of genomic introgression episodes and their role in evolution. Paleogenomic data from species that went extinct during the Early and Middle Pleistocene, such as short-faced hyenas, European jaguars, and the enigmatic *Xenocyon* canids, could shed light on whether these taxa contributed to the genetic make-up of living carnivores. Deep-time paleogenomes could also identify unknown "ghost" lineages that contributed to species' ancestries, as exemplified in the paleogenomic characterization of the Krestovka mammoth (*10*) (Box 1 and Fig. 3). Deep-time DNA can reveal genomic snapshots of an individual species' entire evolutionary story (Box 1). As many temperate and cold-adapted birds and mammals trace their origin to the Early and Middle Pleistocene (*33*, *34*), paleogenomes from these species could correlate evolutionary changes to specific environmental perturbations, such as transitions between climate regimes or community reshuffling. The process of speciation can be investigated as it happens, through exploration of founder event bottlenecks and testing whether speciation occurred through strict allopatry or gradually with post-divergence gene flow. As deep-time paleogenomes tend to occupy basal phylogenetic positions within their clades, they can also provide important calibrations for estimating rates of molecular evolution. For example, paleogenomic data from a Middle Pleistocene hominin from Sima de los Huesos in present-day Spain confirmed hypotheses from Late Pleistocene genomes that Neanderthals and Denisovans diverged during the early Middle Pleistocene (*35*), whereas the inclusion of a ~700-ka horse paleogenome in the equid phylogeny pushed the estimated time for the origin of living equids to more than twice that previously hypothesized (*8*).

Paleogenomes from the Early and Middle Pleistocene can also be used to test hypotheses about relationships among species, including how derived forms are related to earlier forms. An outstanding question in paleontology is whether fossil morphospecies are true species, synchronous ecomorphs, or chronospecies that were direct ancestors of succeeding species. A paleogenomic study of ancient North American bison dating from ~130 to 110 ka, for example, showed that two samples exhibiting extreme size dimorphism and representing supposedly distinct species—the longhorn bison

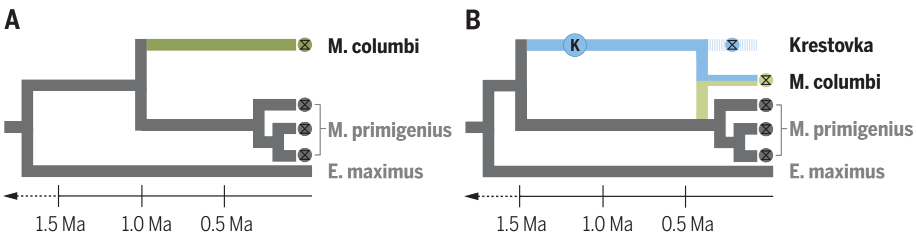


Fig. 3. Deep-time paleogenomes provided new understanding of the evolutionary history of mammoths. Paleontological hypotheses assumed that the *M. columbi* lineage evolved after early divergence from *M. primigenius* (**A**), however, isolation of a deep-time paleogenome from the Krestovka mammoth (blue circle) revealed that *M. columbi* emerged more recently following admixture between the Krestovka and M. primigenius lineages (**B**). Adapted from (*10*).

Box 1. Deep-time mammoth DNA and the inference of a lineage's entire evolutionary story.

The power of a deep-time genomic approach was showcased in a study by van der Valk and colleagues (*10*) in which genome-wide data was collected from three Siberian mammoths dated to approximately 700 ka to 1.2 Ma, which made it possible to examine mammoth evolution from multiple genomic angles:

A new lineage: The oldest of the mammoth specimens belonged to a previously unknown and divergent evolutionary lineage, Krestovka. This implied that two distinct lineages of mammoth, Krestovka and the ancestors of woolly mammoths, lived in Siberia during the later stages of the Early Pleistocene. The analysis also revealed that mammoths belonging to the Krestovka lineage were the first mammoths to colonize North America 1.5 to 1.2 Ma (*61*).

Hybridization: Multiple lines of evidence suggested that the Columbian mammoth originated as a result of hybridization between the Krestovka lineage and early woolly mammoths. This hybridization took place as woolly mammoths expanded into North America during the Middle Pleistocene, after the Krestovka lineage was already established on the continent (Fig. 3). Columbian mammoths derive approximately 50% ancestry from each of these two lineages.

Adaptive evolution: The deep-time nature of the mammoth data set allowed van der Valk *et al.* to estimate the rate of adaptive evolution in mammoths. They concluded that the evolutionary origin of the woolly mammoth lineage did not coincide with an increased rate of protein-coding changes and therefore higher rates of positive selection across the genome (*10*). Subsequent analyses identified a suite of genes that underwent protein-coding changes during the last 700 ka and were thus specific to woolly mammoths (*37*).

and the steppe bison—actually belong to the same lineage that dispersed into North America only a few tens of thousands of years earlier (*36*). Conversely, deep-time paleogenomics can also give context to species for which we have only limited remains, such as Denisovans (*35*).

Finally, paleogenomes across deep-time scales will make it possible to explore aspects of adaptive evolution. At the most basic level, deep-time genomes can help identify when adaptive mutations arose. For example, comparative analysis of mammoth paleogenomes ranging from a few thousand to more than a million years old identified genes associated with hair and skin development, fat storage and metabolism, immune system function, and body size that evolved in the woolly mammoth lineage within the last 700 ka (*37*). Paleogenomes will also allow exploration of how the rate of protein coding changes varies over time, such as in conjunction with past changes in climate, as well as to assess when genomic deletions arose and the rate of positive and purifying selection in introgressed genomic regions.

The impact of glacial cycles on biodiversity

Nearly all ancient DNA studies to date have for practical reasons focused on Late Pleistocene or more recent materials (Fig. 1). Thus, our current understanding of evolutionary processes during the Early (2.6 Ma to 780 ka) and Middle (780 to 126 ka) Pleistocene sub-epochs relies mostly on more traditional approaches, including morphometrics, stable isotope analysis, and pollen records. Increasing access to genomic data from fossils and sediments dating to these earlier sub-epochs will enable more explicit tests of hypotheses about how glacial cycles affect evolution and biodiversity.

A special attribute of the earlier Pleistocene, for example, is the change in periodicity of glaciations from ~40 ka cycles to ~100 ka cycles that occurred 1.2 to 0.7 Ma (*38*) (Fig. 1). This change resulted in isolation of temperate species in glacial refugia for longer periods, providing more time for local adaptation and increasing the rate of population divergence. Biological communities may also have been reshuffled, as the longer and higher amplitude glaciations allowed sufficient ice sheet accumulation for the Bering Land Bridge to form, making land dispersal between Eurasia and North America possible.

Since the change in glacial periodicity, the dominant pattern has been cycles of long glaciations separated by short, warm interglacials. This pattern is believed to have driven the demography and range dynamics of many species (*39*). Long interglacials, for example, have been correlated with bottlenecks in cold-adapted taxa (*40*) and expansion and speciation in warm-adapted taxa (*41*). Of particular interest is the unusually long interglacial that occurred 420 to 370 ka (Marine Isotope Stage 11) (*42*). Paleogenomes from individuals that lived during this long bottleneck and earlier could be used to test these hypotheses and reveal evolutionary changes that were overwritten by subsequent genetic bottlenecks.

Inference of ancient ecosystems

We have described insights that could be derived from DNA extracted from remains of individuals that lived during the Middle Pleistocene and earlier. However, the advances that enable deep-time paleogenomics also make it possible to reconstruct entire deep-time ecological communities. To date, only five studies have attempted to use sedimentary ancient DNA to reconstruct plant and/or animal communities dating to the Middle Pleistocene or older: Kjær *et al.* (*11*) reconstructed components of an Early Pleistocene interglacial ecosystem from sediment extracted from the present day polar desert in northern Greenland; Armbrecht *et al.* (*43*) reconstructed an Early to Middle Pleistocene marine ecosystem from Iceberg Alley in the Southern Ocean; Courtin *et al.* (*12*) reconstructed a Middle Pleistocene interglacial ecosystem from a permafrost megaslump in Eastern Siberia; and Willerslev *et al.* reconstructed Middle Pleistocene plant communities from sediments collected below the Greenland ice sheet (*14*) and from coastal Siberian permafrost (*13*). Among these, Kjær *et al.* and Armbrecht *et al.* enriched libraries for sequences of interest through hybridization to synthesized baits designed to target Arctic or Antarctic taxa. In contrast to the PCR-based metabarcoding approach used by the other three studies, hybridization-based targeted enrichment can capture molecules of any length and are therefore powerful even when preserved molecules are short. Although the hybridization-based approach is currently limited to capturing sequences that are genetically similar to other known taxa, methodological improvements in hybridization capture is a ripe area of research that will expand access to deep-time sedimentary DNA.

Deep-time sedimentary DNA research will enable better understanding of the effect of glacial-interglacial transitions on community composition. Reconstructions of communities spanning the transition into the present Holocene, for example, have revealed rapid biological turnover that closely tracked abiotic changes (*44*, *45*). Comparison with older transitions will test whether patterns are predictable or idiosyncratic, whether some species or communities are more resilient to environmental upheaval than others, and whether some transitions or events leave lasting signatures on community biodiversity.

Reconstructions of communities that thrived in past warm interglacials may provide insight into the potential composition of communities in a future, warmer world (*11*), and improve our understanding of how ecosystem-level interactions among species evolve and are maintained. They also enrich our understanding of these extinct ecosystems beyond what is knowable from the fossil record. Deep-time sedimentary DNA from northern Greenland, for example, revealed a mastodon or mastodon-like animal that was part of the Early Pleistocene community (*11*), although no fossil remains from such an animal have been discovered. Deep-time sedimentary DNA can also reveal past connectivity among populations, as indicated by a recent study of Late Pleistocene sedimentary DNA from a cave in Mexico that linked an extinct population of black bears to living populations in eastern North America (*46*). As technologies improve—particularly those that allow increasingly sensitive targeted enrichment—we envisage deep-time sedimentary DNA as a powerful tool to explore the ecological and evolutionary consequences of environmental change on community-level biodiversity.

Future research to enable recovery of deep-time DNA

It has been shown that DNA can survive in ideal preservation conditions into at least the Early Pleistocene. The next phase of deep-time DNA research is to expand the taxonomic, geographic, and temporal range of recovered and authenticated deep-time DNA. This challenge presents new research opportunities in the field, at the bench, and bioinformatically.

Deep-time genomics is today mostly conducted on substrates with optimal DNA preservation such as those derived from permafrost or caves. However, more efficient approaches to recover ancient DNA molecules will continue to expand the range of samples and substrates suitable for analysis. Today, methods for DNA extraction and library conversion do not recover all potentially preserved DNA molecules. For example, Kjaer *et al.* (*11*) found that DNA adsorbed preferentially to clay mineral surfaces compared with nonclay surfaces, particularly to the clay mineral smectite, which can bind 200 times more DNA than quartz and is a common mineral in terrestrial samples. Their best performing extraction protocol recovered 40% of DNA bound to quartz and only 5% of DNA bound to smectite, suggesting that most DNA was inaccessible. Although anecdotal, this observation points to several opportunities for improving deep-time DNA research, including using mineralogical characterization to identify the most promising sites for deep-time sedimentary DNA recovery and refining experimental approaches to recover DNA bound to all mineral surfaces. In the absence of improved methods to release bound DNA, microscopic evaluation of sedimentary samples will improve the efficiency of DNA recovery. Massilani *et al.*

(*47*), for example, showed that DNA preserved in cave sediment is concentrated in microscale particles, especially fragments of bone and feces preserved within the substrate.

Library conversion protocols could also be made more efficient. Optimized library conversion protocols use enzymatic ligation and polymerization, but ancient DNA extracts contain inhibitors as well as molecules with uncharacterized DNA damage. Although we can convert as little as 100 picograms of DNA into libraries using the Santa Cruz method (*25*), library preparation has been shown to typically convert only ~10 to 50% of extracted DNA (*21*), suggesting that most recovered molecules are lost at this experimental step. Improvements in library preparation may include engineering more robust enzymes to combat inhibitors or developing protocols that incorporate enzymatic repair during library conversion. Additionally, reducing reliance on ligase and polymerase steps through alternative enzymatic strategies, bio-orthogonal chemistry, or native DNA sequencing may offer new approaches to convert currently unsequenceable DNA molecules.

Many species that are obvious targets for deep-time DNA research are extinct and some such as *Xenocyon* canids and basal members of the elephant and horse families have no evolutionarily close living relative for which an ideal reference genome can be produced. This presents challenges to ancient DNA authentication and identification as well as to reference-guided genome assembly. Although the average fragment length of deep-time DNA sequences is short, it may be possible to generate de novo assemblies from ancient extracts by capitalizing on methods that use chromosome conformation capture to retain proximity information useful to link short reads within a chromosome (*48*). Approaches that sequence DNA in situ (*49*) are also promising but are currently in early stages of development. Improvements in bioinformatic processing will also benefit eukaryotic paleogenomic reconstruction and variant calling. Recently, microbial genomes were assembled from DNA recovered from relatively recent paleofecal samples (*50*) and from archaeological dental calculus dating to as old as 100 ka (*51*), suggesting a bioinformatic path toward de novo assembly of some small paleogenomes. Although this approach is not likely to apply to complex eukaryotic genomes, other bioinformatic approaches can improve the accuracy of these assemblies from short read data. Replacing linear single-species reference genomes with multispecies variation graphs that incorporate variants from several genomes (*52*), for example, can increase the number of reads that map to a reference genome. This approach has the additional benefit of allowing variation among indel lengths as well as among nucleotides. Iterative assembly approaches, such as the mapping-iterative-assembler used to generate the first Neanderthal mitochondrial genome (*53*) may improve mapping to more complex genomes. Finally, as reference-based taxonomic assignment is always limited to sequences deposited in public databases, the ongoing population of these databases will continue to improve robust identification of DNA recovered from Early and Middle Pleistocene remains and sediments.

One considerable challenge for studies of deep-time DNA is knowing the age of samples so that they can be placed into broader evolutionary and geological contexts. As most ancient DNA to date is from organisms that lived within the last several tens of thousands of years, it is usually possible to estimate their age directly using radiocarbon dating. However, the short radioactive half-life of carbon-14 means that age estimates are often unreliable if organisms lived more than ~50 ka. Trapped charge dating methods, such as electron spin resonance (ESR) for tooth enamel or luminescence approaches for minerals such as quartz and feldspar, can provide age estimates for samples dating throughout the Pleistocene but require that sediments remain undisturbed since burial [for a review see (*54*)]. When proteins are preserved, the extent of amino acid racemization, hydrolysis, and decay can also estimate time since death, although amino acid "clocks" are known to vary among species and localities (*54*).

In some cases, paleoenvironmental, geological, and geophysical markers can provide clues about a sample's age. A fossil might be found in the Arctic with other paleoecological proxies that suggest a warm and wet environment, for example, indicating that the animal lived during a previous interglacial or in sediments with reversed polarity suggesting that it lived prior to the last paleomagnetic reversal some 780 ka. In some environments, tephra beds—layers of fine, settled volcanic ash—can be dated by methods such as glass fission-track or argon-argon dating. Tephra beds, which are detectable even when present in only microscopic amounts (*55*), have been particularly important in dating sediment cores but can also provide contextual clues about the age of samples found in situ at sites where tephra is present. As volcanic eruptions were common throughout the Pleistocene, improved tephrochronologies spanning the Early and Middle Pleistocene will help place deep-time DNA into a chronological context.

Other approaches to dating deep-time genomes might rely on the predictable nature of evolutionary change in organisms. Molecular clock methods infer the age of paleogenomes by estimating the amount of "missing" evolution along a phylogenetic branch leading to the paleogenome, often called "branch shortening" (*56*). Because the accumulation of mutations is approximately constant over time, the differences between these branch lengths should correspond to the number of generations that separate the represented paleogenome from extant or more recent individuals. To translate missing generations into calendar time, however, the branch shortening approach requires either an independent fossil calibration or an estimate of generation length. For many lineages that lived during the Early and Middle Pleistocene, dated ancestral fossils are few and, with no close living relatives, estimates of generation time would be imprecise. Variation among evolutionary rates between distantly related lineages may also reduce the power of a comparative molecular dating approach. Nonetheless, development of approaches that use genomic information to estimate the age of paleogenomes and their evolutionary relationships to other species is a rich area for future research.

Conclusions

The next decade will bring continued technical advances that will expand the taxonomic and geographic range of deep-time paleogenomes and deep-time ancient sedimentary DNA data sets. These will include new insights into what substrates are likely to preserve deep-time ancient DNA, as well as refined approaches to release DNA bound to biological or mineralogical matrices. These newly assembled deep-time paleogenomes will be placed into chronological context through developments in geochronology and paleoecology, together with more powerful computational approaches to estimate the age of samples using a molecular clock. The resulting deep-time data sets will enable reconstruction of evolutionary histories across repeated environmental perturbations, refining understanding of adaptive evolution, community organization, and ecosystem resilience. Moreover, as the past by its nature is different from anything that exists today, access to deep-time DNA provides ample yet unpredictable opportunities for scientific discovery.

REFERENCES AND NOTES

1. J. Ehlers, P. L. Gibbard, P. D. Hughes, *Quaternary Glaciations - Extent and Chronology: A Closer Look* (Elsevier, 2011).
2. C. Posth *et al.*, *Curr. Biol.* **26**, 827–833 (2016).
3. M. Baca *et al.*, *Proc. Biol. Sci.* **290**, 20222238 (2023).
4. L. Loog *et al.*, *Mol. Ecol.* **29**, 1596–1610 (2020).
5. R. E. Green *et al.*, *Science* **328**, 710–722 (2010).
6. A. D. Foote *et al.*, *Nat. Commun.* **4**, 1677 (2013).
7. C. Valdiosera *et al.*, *Biol. Lett.* **2**, 601–603 (2006).
8. L. Orlando *et al.*, *Nature* **499**, 74–78 (2013).
9. A. Barlow *et al.*, *Curr. Biol.* **31**, 1771–1779 (2021).
10. T. van der Valk *et al.*, *Nature* **591**, 265–269 (2021).
11. K. H. Kjær *et al.*, *Nature* **612**, 283–291 (2022).
12. J. Courtin *et al.*, *Environ. DNA* **4**, 1265–1283 (2022).
13. E. Willerslev *et al.*, *Science* **300**, 791–795 (2003).
14. E. Willerslev *et al.*, *Science* **317**, 111–114 (2007).
15. M. Meyer *et al.*, *Nature* **505**, 403–406 (2014).
16. E. I. Zavala *et al.*, *Nature* **595**, 399–403 (2021).
17. T. Lindahl, *Nature* **362**, 709–715 (1993).
18. J. S. Pedersen *et al.*, *Genome Res.* **24**, 454–466 (2014).
19. A. J. Hansen *et al.*, *Genetics* **173**, 1175–1179 (2006).

20. A. W. Briggs *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14616–14621 (2007).
21. L. Bokelmann, I. Glocke, M. Meyer, *Genome Res.* **30**, 1449–1457 (2020).
22. J. Dabney, M. Meyer, S. Pääbo, *Cold Spring Harb. Perspect. Biol.* **5**, a012567 (2013).
23. E. Hempel *et al*., *Mol. Biol. Evol.* **39**, msac241 (2022).
24. M.-T. Gansauge *et al*., *Nucleic Acids Res.* **45**, e79–e79 (2017).
25. J. D. Kapp, R. E. Green, B. Shapiro, *J. Hered.* **112**, 241–249 (2021).
26. A. W. Briggs *et al*., *Nucleic Acids Res.* **38**, e87 (2010).
27. C. de Filippo, M. Meyer, K. Prüfer, *BMC Biol.* **16**, 121 (2018).
28. K. Prüfer, *Bioinformatics* **34**, 4165–4171 (2018).
29. S. Peyrégne *et al*., *Sci. Adv.* **5**, eaaw5873 (2019).
30. W. Miller *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2382–E2390 (2012).
31. J. A. Cahill *et al*., *Mol. Biol. Evol.* **35**, 1120–1129 (2018).
32. M.-S. Wang *et al*., *Nat. Ecol. Evol.* **6**, 936–944 (2022).
33. J. T. Weir, D. Schluter, *Proc. Biol. Sci.* **271**, 1881–1887 (2004).
34. A. M. Lister, *Philos. Trans. R. Soc. B*, 221–241 (2004).
35. M. Meyer *et al*., *Nature* **531**, 504–507 (2016).
36. D. Froese *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **114**, 3457–3462 (2017).
37. D. Díez-Del-Molino *et al*., *Curr. Biol.* **33**, 1753–1764.e4 (2023).
38. P. U. Clark *et al*., *Quat. Sci. Rev.* **25**, 3150–3184 (2006).
39. J. R. Stewart, A. M. Lister, I. Barnes, L. Dalén, *Proc. Biol. Sci.* **277**, 661–671 (2010).
40. E. Lord *et al*., *BMC Ecol. Evol.* **22**, 126 (2022).
41. J. Ortego, L. L. Knowles, *Mol. Ecol.* **31**, 296–312 (2022).
42. L. E. Lisiecki, M. E. Raymo, L. E. Lisiecki, M. E. Raymo, *Paleoceanography* **20**, (2005).
43. L. Armbrecht *et al*., *Nat. Commun.* **13**, 5787 (2022).
44. A. J. Monteath *et al*., *Quat. Sci. Rev.* **299**, 107878 (2023).
45. C. L. Clarke *et al*., *Sci. Rep.* **9**, 19613 (2019).
46. M. W. Pedersen *et al*., *Curr. Biol.* **31**, 2728–2736.e8 (2021).
47. D. Massilani *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2113666118 (2022).
48. N. Kaplan, J. Dekker, *Nat. Biotechnol.* **31**, 1143–1147 (2013).
49. A. C. Payne *et al*., *Science* **371**, eaay3446 (2021).
50. M. C. Wibowo *et al*., *Nature* **594**, 234–239 (2021).
51. M. Klapper *et al*., *Science* **380**, 619–624 (2023).
52. R. Martiniano, E. Garrison, E. R. Jones, A. Manica, R. Durbin, *Genome Biol.* **21**, 250 (2020).
53. R. E. Green *et al*., *Cell* **134**, 416–426 (2008).
54. K. E. H. Penkman *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2109324119 (2022).
55. S. M. Davies, *J. Quat. Sci.* **30**, 114–130 (2015).
56. M. Meyer *et al*., *Science* **338**, 222–226 (2012).
57. J. A. Fellows Yates *et al*., *Sci. Data* **8**, 31 (2021).
58. J. Von Eggers *et al*., Inventory of ancient environmental DNA from sedimentary archives: locations, methods, and target taxa, Version 1.0, Zenodo (2022); https://zenodo.org/record/6847522.
59. D. Diez-del-Molino, P. D. Heintzman, L. Dalén, A list of representative paleogenomic datasets derived from human and faunal remains, Version 1.0, Zenodo (2023); https://zenodo.org/record/8270285.
60. J. Dabney *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15758–15763 (2013).
61. A. M. Lister, *Quat. Int.* **443**, 14–31 (2017).

ACKNOWLEDGMENTS

We thank D. Diez-del-Molino for assembling the database of paleogenomes used in Fig. 1. We also thank the editor and two reviewers for helpful comments that improved the manuscript. **Funding:** P.D.H. and L.D. were supported by the Knut and Alice Wallenberg Foundation (KAW 2021.0048 to P.D.H. and K.A.W. 2022.0033 to P.D.H. and L.D.). L.D. also acknowledges support from the Swedish Research Council (2021-00625) and the European Union (ERC, PrimiGenomes, 101054984). B.S. was supported in part by NSF OPP-2131589. J.K. was supported by NIJ 2020-DQ-BX-0014. **Competing interests:** Authors declare that they have no competing interests. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

Submitted 31 May 2023; accepted 5 September 2023
10.1126/science.adh7943

EDITORIAL

Modern government and science advice

Governments need to understand science. This is obvious when thinking about defense and security, health, or the challenges of climate change and biodiversity loss, but it is true for all areas of government activity. Science has something to offer in developing policy for town planning, education, transportation, food, environmental management, the administration of justice, communication systems, the use of artificial intelligence to improve public services, and much more. Seven out of the 10 largest companies in the world are based on science and technology, and there is a correlation between spending on R&D and productivity at a national level. Yet in most democracies, scientists make up a tiny fraction of politicians, and recently there have been questions about the overextension of scientific authority (so-called "scientism") or the inappropriate empowerment of scientists. What, then, is the role of a science adviser to modern government?

In 1974, Solly Zuckerman, the United Kingdom's first official scientific adviser, said that decision-making rested with government ministers, and that "If scientists want more than this then they'd better become politicians." This notion was also captured by Winston Churchill, who said "scientists should be on tap not on top." These ideas may be true in a democracy, but it doesn't mean that science won't be ignored or misused by politicians, nor does it mean that what scientists say will be understood.

> **"Science advice requires dialogue, not data dumping."**

Effective interaction with government is challenging for science advisers in the presence of a rising tide of public anti-science sentiment. To overcome mistrust of science, science advisers should consider four key issues when engaging with governments. It is essential to ensure that the evidence base for recommendations is adequate, and if not, figure out what, if anything, can be done to supplement it in a timely manner. It is crucial that advice given to government is understood, including the uncertainties. A major obstacle is that nonscientists may often view science as definitive and providing an immutable truth, rather than being a self-correcting process of discovery. Scientists have not done a good job in describing the process of science, and how consensus can change with time. This has led to the exploitation and subversion of evidence by those who stoke mistrust in science. Wearing masks and vaccination during the pandemic are stark examples.

Advice should be in a form that is relevant to policymakers. For example, although it is interesting to know the precise rate of mutation in coronaviruses, this information is unlikely to help guide policy decisions unless placed into context. What are the implications that follow from that observation? What are the relevant social and behavioral issues? Science advice requires dialogue, not data dumping. Advice to government also must explain how science can be used to monitor the effects of any policy choice. Unfortunately, rigorous evaluation of the effects of new policies, and the transparency of the evaluations, is underdeveloped in many countries.

Effective science advice is not possible unless there is a strong independent science base that government advisers can call on. And it is not just academics that are needed—often, the expertise lies in industry or in government laboratories. Science advice should be published so that the public and the scientific community can provide constructive critique. This is, ultimately, the very way in which science progresses. Publication also shows how and whether science affected the ultimate policy choice made.

Science advice is unlikely to be effective if governments lack scientists. This absence often means that there are not effective mechanisms to absorb recommendations and evaluate policy implications. The lack of scientifically trained individuals in government is striking. For example, in 2018, university graduates entering the "fast stream track" of the civil service in the United Kingdom were overwhelmingly made up of students from the arts and humanities; only about 10% held a degree in science. Attracting more scientists into government will require new approaches to recruitment and training.

Policies must balance conflicting interests while addressing specific problems. Science advice cannot be followed slavishly, but if decision-makers ignore such advice, or treat it as a box-ticking exercise, then the systems that uphold societal well-being will be at risk. The most stark example of this is climate change—and we are witnessing the results of that crisis.

–Patrick Vallance



Patrick Vallance is chair of the Natural History Museum in London, UK. He was the UK Government Chief Scientific Adviser from 2018 to 2023. patrick.vallance@nhm.ac.uk

PHOTO: NATIONAL HISTORY MUSEUM, LONDON, UK

10.1126/science.adl0894

NEWS

IN BRIEF

Edited by **Kelly Servick**



PHYSICS

Fusion experiment beats previous energy record

The National Ignition Facility's reaction chamber (central blue sphere), surrounded by laser beam tubes and diagnostic instruments.

Lightning has struck a second time for physicists using lasers to achieve nuclear fusion, in which two atomic nuclei combine into one while releasing enormous amounts of energy. On 30 July, the 192 lasers of the stadium-size National Ignition Facility (NIF) at Lawrence Livermore National Laboratory instantaneously crushed a tiny capsule filled with heavy isotopes of hydrogen. In doing so, they prompted a fusion reaction that produced more energy than the laser beams deposited onto the target. The new results, presented by NIF scientists at a conference last week, mark the second time in 8 months that NIF has achieved "energy gain." NIF's December 2022 experiment released 3.15 megajoules (MJ) of fusion energy—the equivalent of about three sticks of dynamite, and 54% more energy than went into the target. In the July test, a shot yielded 3.88 MJ, an 89% surplus. The results will help researchers learn which factors separate powerful blasts from duds—a possible step toward one day making fusion a reliable energy source.

Flu shot to drop vanished strain

INFECTIOUS DISEASE | An influenza virus variant that circulated widely before the COVID-19 pandemic has disappeared, leading an expert group that advises the World Health Organization (WHO) to recommend last week that it no longer be included in flu vaccines. The strain, known as Yamagata, has not been detected since March 2020, and some scientists suspect it went extinct because of social distancing measures imposed during the pandemic. WHO's advisory committee, which meets every February and September to determine which viral strains should be in flu vaccines, noted that because making flu vaccines requires growing the virus, a lab accident could in theory reintroduce Yamagata. To reduce this risk, "every effort should be made to exclude it as soon as possible," it concluded. National regulators ultimately decide on the composition of flu vaccines in their own countries.

Lab-developed tests get scrutiny

BIOMEDICINE | The U.S. Food and Drug Administration (FDA) wants to start to regulate a slew of medical tests that now lack stringent oversight. These "lab-developed tests" are made and used within a single lab and include everything from predictors for Alzheimer's disease to genetic analyses of tumors, but they have faced concerns about their validity. Under the new rule, which FDA proposed last

week, the testmakers would be required to submit data on their tests' accuracy as medical device–makers do now. In a statement, FDA noted concerns over misdiagnoses and patients being over- or undertreated for common conditions such as heart disease. Commercial labs and academic medical centers that develop the tests argue that additional regulation will delay scientific advances. FDA will accept public comments on the rule until 4 December.

Good reviews for *eLife* model

PUBLISHING | An analysis shows authors sharing their work in the nonprofit journal *eLife* report largely favorable experiences with the new publishing model it introduced in January, under which it posts manuscripts and associated peer reviews but no decision to accept or reject them. The selective, open-access journal, which focuses on the life sciences, aimed to boost the speed and transparency of publication. Authors can opt for the old model, but more than 90% of manuscripts submitted from February to July went through the new process, says the report, commissioned by *eLife* from an outside firm and released last week. The experiment has been closely watched, with some critics predicting it would drive away quality submissions. According to the report, total submissions declined by one-quarter, to 3223, compared with the same span of months in 2022. (*eLife* said it nearly reached its target.) Reviews were published within 79 days, on average, of the preprint's posting, down from an average of 232 days for publication of accepted papers under the old model. Seventy-one percent of reviewing editors said the quality of manuscripts stayed the same or improved, while only 42% of senior editors agreed.

Biobanking endangered species

CONSERVATION | A new partnership announced this week will preserve the cells of one-quarter of all threatened and endangered mammal species in the United States to aid in current and future recovery efforts. To date, scientists have cryopreserved tissue for only 14% of the approximately 1700 U.S. species listed as threatened or endangered. So the nonprofit Revive & Restore is funding a project with several partners to develop a biobank pipeline. In a pilot, biologists from the U.S. Fish and Wildlife Service are collecting tissue samples from 24 endangered mammal species, including



A Preble's meadow jumping mouse captured in a population survey provided skin cells for biobanking.

the Preble's meadow jumping mouse, the Mexican wolf, and the Sonoran pronghorn. The San Diego Zoo Wildlife Alliance and the cloning company ViaGen Pets & Equine will sequence the genomes and create cell lines, which will be preserved in the U.S. Department of Agriculture's national repository for cryopreservation. Such specimens could be used to clone endangered species and boost the genetic diversity of wild populations.

Grizzly reintroduction planned

ECOSYSTEMS | The U.S. National Park Service and Fish and Wildlife Service are proposing to reintroduce a top predator to the North Cascades mountain range in Washington state. A draft plan, released last week, envisions establishing a population of 25 grizzly bears within 5 to 10 years, with the population then naturally increasing to about 200. Grizzlies live in nearby mountain ranges in British Columbia, Idaho, and Washington, but highways and railroads prevent these bears from dispersing into the North Cascades. According to the plan, bears would be captured from these populations and relocated by helicopter. Ranchers often oppose the recovery of grizzlies and other predators, so the plan lays out options for reducing conflict with ranchers and hikers, such as allowing the bears to be relocated or killed in certain circumstances. The plan is open for public comment until 13 November.

Trust shapes views on science

PUBLIC OPINION | Just 51% of Americans who express low trust in government and other institutions are fully vaccinated for COVID-19 versus 89% with high institutional trust, a survey has found. Although Republican respondents expressed less of this trust than Democrats overall, individuals with high trust were more likely to get vaccinated than those with low trust even within each party, according to results from the online survey, released last week by the right-leaning American Enterprise Institute. Overall, trust in scientists fell during the pandemic: Only 69% of the 5055 respondents expressed confidence that scientists would act in the public's best interests, down from 86% in January 2019.



DEVELOPMENTAL BIOLOGY

How hammerhead sharks get their hammers

With their broad, flattened snout tipped at each end with giant google eyes, hammerhead sharks are both charismatic and easy to spot. But unlike most fish, hammerheads give birth to live young, so researchers have not been able to learn much about the development of any of the 10 or so species. Now for the first time, scientists have captured step by step how this "hammer" forms in developing embryos. Researchers at the University of Florida collected, imaged, and stained embryos of various stages from a small hammerhead species, the bonnethead shark, which were being studied for other reasons. Even though the shark's head forms early on, the hammer doesn't begin to appear until about 2 months later, about halfway through gestation, the team reported last week *Developmental Dynamics*. The hammers of other sharks in this group likely follow the same developmental pathway, the researchers say. Next up: figuring out the genes involved in creating this unusual snout.

A bonnethead shark embryo roughly 3 months into its 4- to 5-month gestation.

PHOTOS: (TOP TO BOTTOM) KIKA TUFF/REVIVE & RESTORE; GARETH FRASER AND STEVEN BYRUM/UNIVERSITY OF FLORIDA

IN DEPTH



A health worker examines a young participant in a clinical trial of the R21 malaria vaccine in Kenya.

GLOBAL HEALTH

Second malaria vaccine gets WHO green light

New shots could make malaria protection more plentiful, saving tens of thousands of lives

By **Gretchen Vogel**

The world will soon have another powerful weapon against malaria, a disease that kills half a million people each year, most of them young children in sub-Saharan Africa. Data from a trial involving 4800 children in four African countries suggest that a vaccine developed at the University of Oxford, known as R21/Matrix-M, provides significant protection against the disease. The results were posted as a preprint hosted by *The Lancet* last week, and on 2 October, the World Health Organization (WHO) recommended R21 to prevent malaria in children living where the disease is endemic.

The decision makes R21 the second WHO-approved vaccine for use against malaria. The first, called RTS,S/AS01 or Mosquirix and made by GSK, was recommended for use in 2021 and has been given to 1.8 million children in Ghana, Malawi, and Kenya. Doses should arrive in nine more countries in Africa by the end of the year. But the 18 million doses of that vaccine expected to be available between now and 2025 "are only about 10% of what we need" to protect the estimated 40 million children who are born every year in malaria-affected areas, says Matthew Laurens, a malaria vaccine expert at the University of Maryland School of Medicine.

R21 could soon be available in much larger quantities. It "has the potential to save tens of thousands of lives each year," Hanna Nohynek, chair of WHO's Strategic Advisory Group of Experts on Immunization, said in a press conference announcing WHO's recommendation.

Like RTS,S, R21 requires multiple initial doses and a later booster. Both vaccines induce immunity with a protein called CSP from *Plasmodium falciparum*, the parasite that causes the deadliest form of malaria. The new vaccine contains a different immune-boosting agent, or adjuvant, and its viruslike particles are more thickly studded with the CSP protein. It therefore requires a lower dose, which means it can be made more easily in larger quantities, says Adrian Hill, a vaccine expert at Oxford who led development of R21.

The Serum Institute of India, one of the world's biggest vaccinemakers, will manufacture R21 and says it can produce more than 100 million doses of the vaccine per year starting in 2024. That would be "a leap forward" for efforts to fight malaria, Laurens says. "Having a second product would be fabulous." And at an estimated price of $2 to $4 per dose, the new vaccine will likely be less expensive than RTS,S, which cost nearly $10 per dose in the first major purchase of the vaccine by UNICEF.

The phase 3 R21 trial included children between the ages of 5 months and 36 months at sites in Mali, Burkina Faso, Kenya, and Tanzania. Most were given prophylactic treatment with antimalarial drugs, also known as seasonal malaria chemoprevention, which is now standard in many malaria-affected regions. Children then received either R21 or a rabies vaccine as a control. At the three trial sites where malaria transmission is limited to the rainy season, the vaccine had 75% efficacy against clinical malaria over 18 months. At the two sites with year-round transmission, the efficacy was 68% after 12 months.

R21 efficacy was slightly higher in children who received their vaccines early, between the ages of 5 and 17 months. That could be a sign that the vaccine is less effective in people who have already been exposed to malaria, Laurens speculates. If so, R21 might have lower efficacy in areas with

very high malaria incidence, where young infants are exposed at an early age, he says.

The vaccine hasn't yet been tested in regions that have high transmission year-round, the WHO recommendations note. Effectiveness should be monitored carefully in such settings during vaccine rollout, the agency says. No trials have compared the efficacy of the two vaccines head to head, but so far R21 looks to be "as good, maybe even a little better, but not substantially better," than RTS,S, says Robert Sauerwein, a malaria vaccine expert at Radboud University.

For both vaccines, protection fades over time. In the R21 trial, children received three doses 4 weeks apart and a booster dose roughly 12 months after the third dose. It's not clear how long R21's protection might last, and ongoing trials are testing the efficacy of further boosters. Alassane Dicko, a malaria expert at the University of Science Techniques and Technologies of Bamako, who helped lead the R21 trial centers in Mali, says there are some indications that a booster every 2 years might be sufficient.

R21 still needs to complete WHO's so-called prequalification process, which assesses not only safety and efficacy, but also manufacturing practices and quality control. That process is underway and is expected to be completed in "the near future," Kate O'Brien, director of WHO's Department of Immunization, Vaccines and Biologicals said in the press conference. WHO's experts predicted that R21 could be widely available for use in the second half of 2024.

Which vaccines countries choose could depend on several things, O'Brien noted. R21 may be less costly, at least at first, but much more is known about the safety and efficacy of RTS,S. So far, few serious side effects of R21 have emerged, but the phase 3 trial is too small to detect rare adverse events.

At least 28 countries have plans to introduce vaccines as part of their malaria control programs, WHO says. The agency's experts emphasized, however, that the vaccine should be used in combination with already established malaria tools, including insecticide-treated bed nets, rapid diagnosis and treatment, and antimalarial drugs to prevent infections.

"It will be really important not to substitute the one intervention for the other," says WHO epidemiologist Mary Hamel, who has helped guide the introduction of RTS,S in Malawi, Kenya, and Ghana. "We don't want to roll out the vaccine and pull back on bed nets, which are also lifesaving," she told the press conference. "Adding these interventions on top of each other is where we're really going to get high impact." ■

With reporting by Meredith Wadman.

CONSERVATION BIOLOGY

Monarch butterfly is not endangered, new review finds

IUCN shifts status to "vulnerable" after analysis challenged

By **Dennis Normile**

In an unusual reversal, the International Union for Conservation of Nature (IUCN) has decided North America's monarch butterfly is not "endangered." Instead, the insect is only "vulnerable" to extinction, the group said last week—adding that it could lower the alarm still further, changing the listing to "near threatened" if an upcoming census suggests the population is stable or growing.

The 27 September decision followed a researcher's challenge to population models an IUCN team used to justify the endangered designation, conferred just 14 months ago. The team committed a "scientific injustice" by ignoring data showing monarchs are "doing really well," argued ecologist Andrew Davis of the University of Georgia.

IUCN's shift marks the latest twist in a scientific debate over the health of the showy black and orange insect. Monarchs are found worldwide, but the North American subspecies, called the migratory monarch (*Danaus plexippus plexippus*), has become "a poster child of species conservation because of its awesome ecology and migration," says ecologist Anurag Agrawal of Cornell University, who was not involved in IUCN's assessment or the challenge.

Migratory monarchs are split into two populations separated by the Rocky Mountains. Each fall, butterflies in the larger eastern population make an epic migration of up to 4000 kilometers from Canada and the northern United States to a forest in central Mexico, where they overwinter. Then, their descendants make a stepwise return to the north, taking up to four generations to complete the journey. The western population winters in Southern California along the Pacific coast and breeds along the Rockies.

Researchers generally agree that the number of monarchs wintering in Mexico declined beginning in the 1990s or earlier but stabilized around 2014 at about 55 million individuals. Many researchers blame the decline on two factors: logging in Mexico's forests and farming in the U.S. and Canada that increasingly relied on crops modified to resist herbicides. That increased the use of herbicides that kill milkweed, which monarch caterpillars feed on.



Each winter, monarch butterflies migrate up to 4000 kilometers to a forest in Mexico, where they overwinter.

PHOTO: ERICH SCHLEGEL/ZUMA PRESS WIRE

A counterscenario, put forward by Davis and others, is that populations were anomalously high decades ago thanks to large-scale clearing of forests for agriculture in the 1800s. The resulting open landscapes allowed milkweed and monarchs to flourish until the land shifted back to forest or intensive farming. Davis also notes that overwintering populations have crashed in the past only to recover the following summer.

In 2022, an IUCN team released an assessment that used annual counts of the

wintering population from 1993 to 2020, and two population models, to discern long-term trends in monarch numbers, which can vary year to year. One model rested on a "linear" method that assumed a constant rate of change over time. Modeling found that the population could have shrunk by 22% to 72% over 10 years, meeting IUCN's criteria for an endangered designation. The assessment also noted that the stability in overwintering monarch numbers seen since 2014 made "current rates of decline less concerning than they were in years past."

Still, many scientists thought the "endangered" listing was warranted because drought along migration routes or cold winters could tip the population into an extinction spiral. "Monarch populations [are] at a level that most scientists suggest is not sustainable," says Karen Oberhauser, a conservation biologist at the University of Wisconsin–Madison who was on the assessment team.

But last week an IUCN panel ruled that the linear model "cannot be considered plausible." Instead, it suggested a more complex model—published in February 2020 in *Frontiers in Ecology and Evolution* by a team led by Wayne Thogmartin of the U.S. Geological Survey—is better. It shows the eastern monarch population reached an inflection point around 2014, with a steep decline giving way to a slower decline or a slight increase. Such results support a vulnerable designation, the panel said. And if the upcoming winter census shows continuing stability, it said "the most plausible category would become near threatened."

IUCN's ruling is not binding on its member nations, including the U.S., but it can influence policy. In 2020, U.S. officials concluded that legally designating the monarch as a threatened or endangered species was "warranted," but not a priority.

The new designation is drawing mixed reactions. Given the difficulty of forecasting future monarch populations, "the precautionary principle would suggest keeping the listing at endangered, with the potential to uplist" later, says ecologist Leslie Ries of Georgetown University, who was not involved in the assessment or the challenge.

Agrawal, who believes "the original listing of monarchs as endangered was in haste," is less worried. And Davis says he was "actually hoping that [IUCN] would just list [the monarch] as 'least concern,'" its lowest threat level. He is considering submitting his own assessment to IUCN.

In the meantime, Anna Walker, an entomologist at the New Mexico BioPark Society who led the IUCN panel that recommended the endangered listing, notes that the vulnerable listing "still indicates a high level of extinction risk." ■



RESEARCH POLICY

No U.S. shutdown, but dim outlook for science funding

Congress faces strong pressure to hold down 2024 spending

By **Jeffrey Mervis**

U.S. scientists bracing for a government shutdown that would have furloughed federal researchers and disrupted grantmaking are relieved that Congress averted a closure with a temporary spending measure passed last weekend. But Congress is still a long way from approving 2024 spending bills for research agencies. And scientists are likely to be disappointed with many final numbers.

The so-called continuing resolution (CR) passed overwhelmingly on 30 September by both chambers allows agencies to operate until 17 November with spending at current levels. That means science agencies can continue to accept and review grant applications and carry out routine business, although they are forbidden from starting any programs or reshuffling funds.

If lawmakers can't agree on a 2024 spending plan by mid-November, however, they will need to pass another CR to avoid a shutdown. And many observers are betting that a continued lack of consensus will trigger another legislative clash—and possible shutdown—in the days before the holiday season.

For scientists, the bad news is that the chances are dwindling that research agencies will see any of the healthy budget increases that President Joe Biden has requested for the 2024 fiscal year that began on 1 October. And those agencies could fare much worse by the time the dust settles.

One reason is that a late May deal to prevent the government from defaulting on its debt obligations also holds all nondefense discretionary spending—which includes civilian research—at this year's levels rather than the 7% increase Biden has proposed. Defense would grow by 3%, as Biden has requested.

But even that hard-fought compromise between Biden and House of Representatives Speaker Kevin McCarthy (R–CA) is in serious trouble. Prodded by a group of conservative Republicans, the House last week passed four spending bills that would cut the budgets of several civilian agencies below 2022 levels. None covers such key research agencies as the National Institutes of Health (NIH) or the National Science Foundation. But the House is expected to approve similarly sharp cuts for those agencies in coming weeks.

"I would not be surprised to see a bill come to the House floor that contains deep cuts to NIH," says Ellie Dehoney of Research!America, which advocates for biomedical research.

Those low numbers won't be the final word. The Senate has begun to advance its own versions of the 12 spending bills cover-

Another government shutdown looms in 6 weeks if Congress fails to pass a final budget.

ing all government activities; they would give small increases to several research agencies. Those boosts are possible because the Senate is assuming overall spending levels that are much higher than those supported by the House. The two chambers—and the White House—will need to find middle ground before any of the appropriations bills can be passed and signed into law.

Striking a compromise, however, will be complicated by McCarthy's tenuous grip on his position as speaker of the House, where Republicans hold a slim majority. Many of the same group of conservative legislators who supported a shutdown and seek deep budget cuts also want to oust McCarthy as speaker. That leadership drama could complicate progress on spending bills.

"Will McCarthy stick to his agreement with Biden or … push for additional cuts?" asks one science lobbyist who requested anonymity. "I think the answer will depend on his strategy for holding onto his job."

Amid all the uncertainty, science advocates hail the CR as an example of what can happen when congressional Democrats and Republicans focus on getting something done rather than scoring political points. For example, McCarthy dropped proposed language on such divisive issues as abortion and immigration to come up with a bill that Democrats would support. And Senate Majority Leader Chuck Schumer (D–NY) reluctantly removed an additional $6 billion to support Ukraine's military to accommodate some Republicans. The final votes on the "clean" CR—335 to 91 in the House and 88 to nine in the Senate—reflect a willingness to call a truce to the fierce partisanship, says Matt Hourihan of the Federation of American Scientists.

"As a bipartisan measure, the CR further demonstrates that compromise is not only possible, but necessary," Hourihan says. "There's a long road ahead, but I think it sets a good precedent that legislators can build on."

One early test could come if Congress tries to pass a separate bill with Ukraine funding. Pundits predict that winning House support for this "supplemental" could require including funds for border security and restrictions on immigration. Given the consensus on freezing civilian spending, however, a successful supplemental could make it even harder for legislators to boost research budgets.

Such uncertainty makes it difficult to predict when Congress might ultimately agree on a 2024 budget. "I think the next 45 days are going to be very fluid," Dehoney says. "So I'd be wary of anybody offering up any predictions." ■

PHOTO: LAUREN PETRACCA/*THE NEW YORK TIMES*/REDUX

PHYSICS

Third retraction looms for superconductivity physicist

Ranga Dias's co-authors request retraction after doubts cast on key data in blockbuster room-temperature claim

By **Dan Garisto**

Facing a mutiny by his co-authors, Ranga Dias, the University of Rochester (U of R) physicist embroiled in controversy over his superconductivity research and allegations of scientific misconduct, is set to have a third paper retracted.

In the 8 March paper in *Nature*, Dias and his colleagues claimed to have discovered a material that could conduct electricity without resistance at room temperatures and moderate pressures. If the paper is retracted, it will follow retractions of two other claims from Dias's teams: one last month from *Physical Review Letters* (*PRL*), and one in September 2022 of a different *Nature* paper.

On 1 September, *Nature* attached an editor's note to the March paper, warning readers that "the reliability of data presented in this manuscript is currently in question." A week later, eight of the co-authors on the 11-person paper submitted a letter to *Nature* requesting the study be retracted, *The Wall Street Journal* first reported last week. *Science* has obtained the letter and additional documents, which raise concerns about the reliability of the data and Dias's treatment of his co-authors. "We respectfully request and recommend that *Nature* issue a retraction," conclude the signatories, who include five recent graduate students of Dias's.

The documents indicate *Nature* plans to retract the paper. In a 14 September email to the co-authors, *Nature* Senior Editor Tobias Rödel says a postpublication review revealed issues that are "mostly unaddressed." Rödel added, "We are in absolute agreement with your request that the paper be retracted." Responding to *Science*, *Nature*'s Chief Physical Sciences Editor Karl Ziemelis said the journal is "carefully" investigating concerns. "We expect to take action in the near future."

In emails throughout the postpublication review process, Dias reiterated his "unwavering confidence in the integrity of our data." He did not respond to a request for comment. "There are not many new facts in the big picture, but the brazenness and/or incompetence is remarkable," says Peter Armitage, a condensed matter physicist at Johns Hopkins University who reviewed the documents obtained by *Science*.

The March paper drew headlines because of the potential applications for a room-temperature superconductor, such as lossless electrical grids and powerful magnets for trains and medical imaging. In the study, Dias and his colleagues reported that a compound of lutetium, nitrogen, and hydrogen superconducts at temperatures up to 294 K when squeezed to a comparatively moderate pressure of 10,000 atmospheres. (Dias had made earlier claims of creating near–room-temperature superconductors, but only at pressures of more than 1 million atmospheres.)



Ranga Dias is facing another retraction for his work on room-temperature superconductors.

Questions about the paper arose immediately after its publication. On 2 May, two researchers submitted an anonymous critique of the paper to *Nature*. These researchers disclosed their identities to *Science*: They are James Hamlin, a high-pressure experimentalist at the University of Florida, and Brad Ramshaw, an expert in superconductivity at Cornell University.

Hamlin and Ramshaw focused on data supporting the observation of zero electrical resistance, a hallmark of superconductivity. Although the paper presented figures showing the resistance dropping

to about zero, the authors acknowledged modifying the data by subtracting a background signal. Background subtraction is sometimes done in high-pressure superconductivity studies, but the practice is controversial because it can give the false appearance of zero resistance.

In their critique, Hamlin and Ramshaw pointed out that the paper describes a background subtraction method that produces a resistance plot drastically different from the published version. On 28 May, Dias and the other senior author on the paper, Ashkan Salamat, a physicist at the University of Nevada, Las Vegas, provided *Nature* editors with updated instructions on the background subtraction method. But applying this method still produced a figure different from the published plot. Essentially, they could not explain how they arrived at a plot that showed zero resistance, Hamlin says. "The problem is they're not able to answer the question of how much they subtracted off."

In response to the critique, *Nature* initiated a postpublication review process, soliciting feedback from four independent experts. In documents obtained by *Science*, all four referees expressed strong concerns about the credibility of the data. "I fail to understand why the authors … are not willing or able to provide clear and timely responses," wrote one of the anonymous referees. "Without such responses the credibility of the published results are in question." A second referee went further, writing: "I strongly recommend that the article by R. Dias and A. Salamat be retracted."

Most of the paper's co-authors were not included in these initial deliberations until 6 July, at which point Dias provided them with a PDF of the postpublication peer review. In their letter to *Nature* the co-authors allege that they had raised concerns about the study prior to publication, but that Dias dismissed them. They say Dias gave some of them an ultimatum: Remove their names from the paper or allow it to proceed. "Neither choice seemed tenable given that Dr. Dias was in control of our personal, academic, and financial circumstances," they write. "We did not feel that we were able to speak freely."

After seeing the postpublication review, the co-authors used their access to the raw data to corroborate Hamlin and Ramshaw's concerns about the zero-resistance measurement. They found that the paper relied on a method to determine the resistance that was not discussed or disclosed to the co-authors and did not match the raw data. They also argued that another key piece of data presented as evidence of superconductivity, the heat capacity of the sample, was a "flawed measurement."

Before the co-authors sent their letter to *Nature*, Dias sent a cease-and-desist letter to six of them—five of his former graduate students and one U of R faculty member. He warned them about the "potential legal consequences of your actions and to consider the ethical implications of making baseless allegations against a colleague and fellow scientist." Dias has previously sent cease-and-desist letters to other critics of his work. "Raising concerns about data and processes is an integral part of our scientific practice," says Shanti Deemyad, a condensed matter physicist at the University of Utah. "The students are courageous for voicing their concerns and doing the right thing, and the scientific community should protect them."

> **"The students are courageous for voicing their concerns and doing the right thing, and the scientific community should protect them."**
> **Shanti Deemyad**, University of Utah

All co-authors signed the letter to *Nature*, except Dias and two of his U of R students. The signatories also include Salamat, a longtime collaborator who with Dias co-founded a spin-off company, Unearthly Materials, to commercialize superconductors. Earlier this year, during *PRL*'s investigation of its now-retracted paper, Salamat provided what the journal concluded was false raw data in "a deliberate attempt to obstruct the investigation." Salamat did not respond to requests for comment.

Dias is also under investigation by Washington State University for allegedly plagiarizing more than 20% of his Ph.D. thesis. He has previously claimed to be "addressing these issues directly with his thesis adviser." U of R had already conducted two internal investigations of Dias's research that it says found no evidence of misconduct. However, a U of R spokesperson says a third investigation is now underway by external experts who are looking into "the integrity of data across multiple papers."

A source who requested anonymity says that until that investigation has concluded, Dias is not allowed to supervise undergraduate and graduate students. The U of R spokesperson would not confirm whether Dias has also been stripped of his teaching duties. But the university's course website shows that for this fall term, there are zero students enrolled in Dias's classes. ■

Dan Garisto is a science journalist in New York City.

PLANETARY SCIENCE

Astronomers may have spied second Kuiper belt

New objects found beyond Neptune during hunt for targets for NASA spacecraft

By **Paul Voosen**

There just doesn't seem to be enough of the Solar System. Beyond Neptune's orbit lie thousands of small icy objects in the Kuiper belt, with Pluto its most famous resident. But after 50 astronomical units (AU)—50 times the distance between Earth and the Sun—the belt ends suddenly and the number of objects drops to zero. Meanwhile, in other solar systems, similar belts stretch outward across hundreds of AU. It's disquieting, says Wesley Fraser, an astronomer at the National Research Council Canada. "One odd thing about the known Solar System is just how bloody small we are."

A new discovery is challenging that picture. While using ground-based telescopes to hunt for fresh targets for NASA's New Horizons spacecraft, now past Pluto on a course out of the Solar System, Fraser and his colleagues have made a tantalizing, though preliminary, discovery: about a dozen objects that lie beyond 60 AU—nearly as far from Pluto as Pluto is from the Sun. The finding, if real, could suggest that the Kuiper belt either extends much farther than once thought or—given the seeming 10-AU gap between these bodies and the known Kuiper belt—that a "second" belt exists.

The discovery, being prepared for publication and not yet peer reviewed, is supported by measurements from New Horizons itself, which at 57 AU continues to streak beyond the edge of the known Kuiper belt. Many of its instruments are in hibernation, but a dust counter has run continuously during the mission. Dust is a telltale sign of colliding planetary bodies, and so the New Horizons team expected the amount of dust to fall off steeply after the probe left the Kuiper belt, where it had rendezvoused with an object called Arrokoth. Instead, "The number of impacts is not declining," says Alan Stern, the mission's principal investigator and a planetary sci-



A dozen distant objects spotted by the Subaru Telescope in Hawaii are challenging theorists.

entist at the Southwest Research Institute. "And the simplest explanation for that is that there is more stuff out there that we haven't detected."

Astronomers outside the team are intrigued but want to see more evidence. "If there really is a new belt, that's a super-exciting thing," says Pedro Bernardinelli of the University of Washington. However, the findings are hard to reconcile with unpublished results from a recent survey of the outer Solar System using the 4-meter Víctor M. Blanco Telescope in Chile. It surveyed a different slice of sky and turned up only one object beyond 50 AU. Other surveys have also come up dry. "Why are we not seeing these things?" Bernardinelli asks. "Did everyone get unlucky? It's possible, but it's hard."

Stern would welcome more icy bodies to study. After threats to shift New Horizons from planetary science to heliophysics—the study of the Sun and its plasma-filled envelope—NASA relented last week, extending the mission's current focus until the decade's end. That means the probe could still visit another object like Arrokoth, if one exists, Stern says. "We have the opportunity to do science that we know we can do—and can't be done other ways—for a very long time."

For years astronomers have scouted targets for the mission using a wide-field camera on Japan's 8.2-meter Subaru Telescope, on Hawaii's Mauna Kea. The giant telescope is needed, Fraser says, because Kuiper belt objects, so faint and slow moving, "are a pain in the butt to find." Targets for New Horizons are even harder to see because the spacecraft, as seen from Earth, appears to be flying straight toward the bright center of the Milky Way, dazzling instruments.

To boost the signal, the team electronically combines hundreds of images from a night of observation, hoping a bright blob will appear along a likely trajectory. Initially, the team examined the images manually, sifting through 15,000 candidates a night. "You go blind pretty quickly," Fraser says. Now, artificial intelligence makes the screening much less painful. "We went from a week of agony for an army of people to 6 hours of vetting."

The result was the 12 far-out objects. They add to a clue: three blips spotted by star-tracking sensors on the Hubble Space Telescope in its first 20 years, says Hilke Schlichting, an astrophysicist at the University of California, Los Angeles. The sensors, which stare unendingly at single stars to help with pointing the telescope, briefly dimmed for a fraction of a second, suggesting the stars were eclipsed by unknown Kuiper belt objects crossing in front of them. "Is there a larger population beyond 60 AU? There could be," Schlichting says. "Maybe that's what we're seeing. I'm not sure."

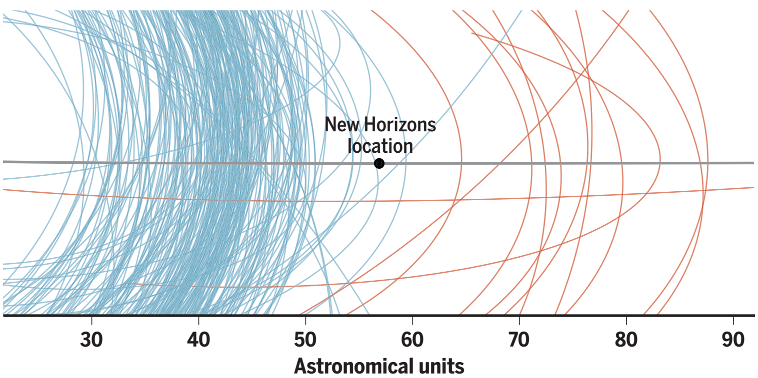
The team has done its best to confirm the objects' orbits, by capturing them at three or four spots along their apparent courses. They think one reason the Blanco telescope hasn't turned up similar results might be because the objects are clustered near New Horizons's path, perhaps by Neptune's gravity, Fraser adds.

Just as intriguing as the new objects is the apparent gap between 50 and 60 AU, says Mihály Horányi, a space physicist at the University of Colorado Boulder who oversees New Horizons's dust counter. "One way or another, something is responsible for maintaining that gap." In other solar systems, planets orbiting within a dusty disk carve gaps by hoovering up material. But no large planet has been seen in the gap. The gap could also be a relic from the Solar System's infancy, caused by waves of pressure in the disk.

The team returned to Subaru for several days last month, using a new optical filter that Fraser says will let them see fainter and smaller objects. They are now analyzing the data. If the second Kuiper belt is real, more than a dozen new distant bodies should emerge, he says. "And if that expectation isn't met, then we have completely missed something." ■

Loosening the belt

Since 2021, astronomers have found hundreds of new objects in the Kuiper belt, a region beyond Neptune. Most sit within 30 and 50 astronomical units (AU, blue). But 12 lie past 60 AU (red)—suggesting NASA's New Horizons mission could have more objects to explore.



CREDITS: (PHOTO) INTERNATIONAL GEMINI OBSERVATORY/NOIRLAB/NSF/AURA/T. SLOVINSKÝ; (GRAPHIC) D. AN-PHAM/SCIENCE; (DATA) SIMON PORTER/SWRI/JHU APL



Katalin Karikó (left) and Drew Weissman's long-te[rm] collaboration paid off in lifesaving vaccines.

NOBEL PRIZES

Award honors pair for mRNA work key to COVID-19 vaccines

Katalin Karikó and Drew Weissman overcame a major roadblock to a novel immunization approach

By **Catherine Offord** *and* **Jon Cohen**

Over the past 3 years, COVID-19 has provided a life-and-death testing ground for a previously unproven vaccine technology. Relying on strands of RNA to rouse the immune system against SARS-CoV-2, the approach resulted in shots that saved countless lives and protected millions from serious illness. Now, Katalin Karikó and Drew Weissman have been awarded this year's Nobel Prize in Physiology or Medicine for messenger RNA (mRNA) research key to the novel COVID-19 vaccines—and their insights could lead to medicines for many other diseases.

The prize, announced this week, recognizes the pair's pioneering studies of mRNA, the genetic strands encoded by DNA to help instruct cells how to assemble proteins. Working together at the University of Pennsylvania in the early 2000s, the two found a way to overcome a major obstacle to using labmade mRNA in medicine: the fierce inflammatory responses it provokes in the body. They identified specific chemical changes to the molecules that tame those responses, paving the way for delivering mRNA into human cells to make viral proteins. Moderna and the Pfizer-BioNTech collaboration both used the approach to quickly make COVID-19 vaccines that began to protect people around the world at the end of 2020.

"This is a wonderful choice of Nobel Prize," says Anthony Fauci, former director of the U.S. National Institute of Allergy and Infectious Diseases, who had Weissman as a fellow in his lab. "[It's] the classic example of a years and years–long collaboration on a very, very difficult problem. … They doggedly kept at it and made a discovery that has already transformed many areas of biomedical research."

Traditional vaccines typically contain a weakened or killed version of a pathogen, key pathogen proteins made in the lab, or viruses genetically engineered to make those proteins in the body. In contrast, mRNA instructs cells to make pathogen proteins that trigger protective immune responses. The approach promised an efficient way to make vaccines, as well as therapeutic proteins—if the potentially harmful immune responses could be forestalled.

Karikó, a biochemist born in Hungary, and Weissman, a U.S.-born immunologist, realized that natural mRNAs, unlike the lab versions, have various chemical modifications that prevent this problem. In 2005, they reported in the journal *Immunity* that replacing an mRNA base called uridine with naturally occurring, modified bases such as pseudouridine could greatly reduce the inflammatory response. They also showed that the change could increase the amount of protein made by cells, generating a strong protective response. Moderna and the Pfizer-BioNTech collaboration used these insights to create mRNA vaccines evoking immunity to SARS-CoV-2 spike protein. (BioNTech employed Karikó until recently.)

It's still unclear whether mRNA vaccines generally offer better protection than other proven vaccine approaches. But they can be faster and easier to produce. Developers simply synthesize the genetic sequence for the desired protein and encase it in a protective lipid nanoparticle. Proteins, viruses, and bacteria, in contrast, need to be produced in bioreactors that are prone to contamination and poor yields.

Karikó and Weissman struggled to find the resources to develop their mRNA work clinically. Karikó couldn't win funding from the National Institutes of Health, and she didn't have a research team when she was doing the key studies, forcing her to do the biochemistry experiments with her own hands. "I am so unlikely a person to win the Nobel Prize," she says.

Karikó, now a professor at the University of Szeged, is just the 13th woman to be awarded the Nobel Prize in Physiology or Medicine in its more than 120-year history. "[Her] grit and determination to follow the science, despite many telling her she could not succeed, are an inspiration to all," says Jessica Foster, an oncologist at the Children's Hospital of Philadelphia.

There's optimism that mRNA medicines will move beyond preventing diseases and into treating them. Companies and academic labs are clinically testing mRNAs that produce therapeutic proteins in the body for conditions ranging from cystic fibrosis to heart disease. Foster, for example, has worked with Karikó on mRNAs that stimulate antitumor immune responses.

Olle Kämpe, vice chair of the committee that chose the winners of this year's award, hopes the honor will increase use of COVID-19 vaccines. The Nobel imprimatur, he suggests, "can make hesitant people take the vaccine and be sure that it's very efficient and it's safe."

Soumya Swaminathan, who served as chief scientist at the World Health Organization until 2022 and now chairs a foundation in Chennai, India, says there is no doubt about the long-term impact of the laureates' work. "The success of mRNA vaccines against COVID has now opened up a whole research enterprise." ■

NOBEL PRIZES

Sculptors of short light pulses win physics Nobel

Researchers share prize for attosecond laser flashes that capture the motion of electrons

By **Adrian Cho** *and* **Daniel Clery**

This year's Nobel Prize in Physics honors three scientists who developed techniques to glimpse the movement of electrons in atoms and molecules using flashes of laser light lasting just attoseconds—billionths of a billionth of a second.

Pierre Agostini of Ohio State University, Ferenc Krausz of the Max Planck Institute of Quantum Optics, and Anne L'Huillier of Lund University share the prize of roughly $1 million for experiments that "enabled the investigation of processes that are so rapid they were previously impossible to follow," the Royal Swedish Academy of Sciences announced Tuesday morning.

"The Nobel committee made the right choice," says Mauro Nisoli, head of the attosecond research center at the Polytechnic University of Milan. "Now, this is a big topic with many active researchers, but these really were the pioneers."

L'Huillier produced such pulses first, following a fortuitous discovery. In 1988, she and her colleagues shined an intense infrared laser through a noble gas and found that the gas emitted ultraviolet light with frequencies that were multiples of the frequency of the original laser light. Over the next several years, L'Huillier and colleagues explained how such "high harmonics" emerge as the laser light rips an electron from a gas atom and then shoves it back in, causing it to emit light.

Those high harmonics are spread across a range of frequencies differing by factors of 30 or more—just what you need to make a very short pulse of light. Fundamentally, any short pulselike wave must contain a spread of frequencies—and the shorter the pulse, the wider the spread. It's the same reason why the information-packed pulses of high-speed internet require "broadband" wavelengths. In the 1990s, L'Huillier and colleagues demonstrated that the high harmonics can be combined in such a way that the waves interfere with each other, producing a train of shorter pulses where the peaks coincide.

Theory predicted that those pulses would be just attoseconds long. But physicists had no way to prove that until Agostini and his colleagues tackled the problem in 2001. He and his team used an argon gas to generate a train of ultrashort pulses and shined both those pulses and part of the original laser pulse that spawned them onto another argon sample. The researchers then took advantage of a prediction, based on interference effects, that the number of electrons emerging at a particular energy would depend on how the two laser pulses overlapped in time. They could vary the overlap by, say, moving a mirror in their optical set up by a micrometer or so. Using this now-standard technique, Agostini and colleagues showed that the pulses were just 250 attoseconds long.

Krausz pushed attosecond pulses further by creating solitary pulses that can be more useful for experiments. In 2001, he and his team produced a lone pulse 650 attoseconds long. To do that, the researchers had to first produce a very wide range of harmonics and then use a filter to select and combine just a few of the highest ones.



The work of Anne L'Huillier, Ferenc Krausz, and Pierre Agostini (left to right) captures fast chemical processes.

PHOTOS: (LEFT TO RIGHT) TT NEWS AGENCY/REUTERS; ALEXANDRA BEIER/GETTY IMAGES; MICHEL EULER/AP

In the past 2 decades, applications of attosecond science have begun to grow. Scientists can measure how long it takes for a pulse to tug an electron away from an atom—a sign of how tightly it is bound—and see how collections of electrons move from place to place inside molecules or materials, key to understanding their properties. Researchers have also used attosecond pulses as ultrafast electronic switches, flipping a material from an insulator to a conductor. Attosecond light pulses may even help diagnose disease in blood samples, by giving molecules a shove and provoking a telltale response.

So far the techniques have produced few major discoveries, says Karl Krushelnick, a physicist at the University of Michigan. "It's kind of a spectacular technology looking for a killer app." But Krushelnick notes that it also took lasers decades to fulfill their initial promise.

Some physicists are working to extend attosecond techniques to other parts of the electromagnetic spectrum. Researchers have now coaxed attosecond pulses out of the Linac Coherent Light Source, an enormous x-ray free-electron laser at SLAC National Accelerator Laboratory, says Linda Young, an atomic physicist at Argonne National Laboratory. In principle, ultrashort x-ray pulses would have enough energy to probe atoms and molecules with much more strongly bound electrons.

L'Huillier is just the fifth woman to win the Nobel Prize in Physics. "L'Huillier was instrumental and for a long time she didn't get the recognition for it," says Ursula Keller, head of the ultrafast laser group at ETH Zürich.

Speaking by phone during the press conference announcing the prize, L'Huillier said there is plenty of fertile terrain left to explore. One goal would be to use attosecond pulses not just to investigate the movement of electrons during chemical processes, but to control them, perhaps driving a reaction that nature doesn't prefer. "The dream—if you want, the Holy Grail—of this part of our research is to be able to control the initial time for a molecular reaction," L'Huillier said. "But this is in the future." ■

With reporting by Phie Jacobs.



FAMILY TIES

Giant family trees based on ancient DNA from thousands of people are revealing prehistoric politics and social structure

By **Andrew Curry**

The burial jar, found under the floor of a mountaintop citadel called La Almoloya in southeastern Spain, held a puzzle. Almost 1 meter in diameter, the vessel entombed a woman in her late 20s with a shining silver diadem on her forehead. She also had silver earplugs threaded through with silver hoops, an awl covered in silver—and a companion: a middle-age man laid to rest in the same jar with a fraction of her wealth. The pair were likely prominent members of a Bronze Age protostate called El Argar, which dominated much of the Iberian Peninsula from hilltop strongholds for nearly 700 years, beginning around 2200 B.C.E.

When archaeologists first excavated these tombs more than a century ago and found many women and men buried together in large jars, they assumed the pairs were royal couples. But given the wealth gap between women and men, later researchers concluded the jars held relatives who died years apart. They thought the women's rich adornments suggested a matriarchy, where powerful women were buried and their sons or grandsons added to their tombs later.

Then, in 2019, a team of geneticists and archaeologists extracted DNA from the La Almoloya woman and her companion, along with 66 other people buried in pairs and singly in the walls and under floors of the hilltop houses. Rather than ancestors and descendants, the analysis showed, the couples were partners. In three cases, children were buried near their parents, who were buried together; one baby girl was the daughter of the woman and man buried in the palace. "We did not expect that," says co-author Roberto Risch, an archaeologist at the Autonomous University of Barcelona. "But [the couples] must have been contemporary, because they had offspring together."

The results also revealed that the elite women seemed to be transplants, perhaps from other high-status hilltop settlements. Unrelated to others buried at the site, the women married men who were related to each other and apparently local, perhaps as a way to knit far-flung El Argar settlements into a cohesive early state. "Probably women were crucial in social networking," Risch says. "It's clear the culture's protagonists are women."

The results from La Almoloya, published last year, are part of a surge of new studies that are shifting the focus of ancient DNA research from genetic links between populations toward intimate, interpersonal connections. As the cost of DNA sequencing has plummeted, researchers have started to sequence genomes from many people at a site, revealing the structure of ancient communities. "It's gotten so cheap that you can do whole cemeteries," says computational biologist Harald Ringbauer of





DNA revealed the connection between a woman and man buried in a jar (right) in the Bronze Age fortress of La Almoloya in Spain (left).

PHOTO: AUTONOMOUS UNIVERSITY OF BARCELONA

the Max Planck Institute for Evolutionary Anthropology (EVA).

Combined with more traditional archaeological data, this tsunami of ancient DNA promises unprecedented insights into prehistoric family practices, group identity, and power. "We can infer not only biological kinship, but social practices," says geneticist Vanessa Villalba-Mouco of EVA. "We can understand marriages, maybe, in the past."

Already, geneticists and archaeologists have teamed up to dissect the marriage patterns of Bronze Age steppe nomads and traced thousands of living descendants of enslaved women buried 200 years ago in a forgotten Maryland cemetery. Forthcoming ancient family trees, presented at two recent conferences and as preprints, promise to reveal not just siblings and parent-child relationships, but distant cousins 10 degrees removed, such as at an Iron Age English site where the burials of 170 relatives spanning 10 generations are grouped by maternal ancestry. At the International Society for Biomolecular Archaeology meeting last month, researchers reported tracing kinship ties among hundreds of people from a remote Polynesian island, showing that millennia-old social practices on the island shifted along with new arrivals.

"There's extraordinarily exciting stuff happening, and it's not just geneticists who are excited about it," says Australia National University archaeologist Catherine Frieman. "We're finally at a point where you can actually ask archaeological questions about this data," like how people in the past organized societies and venerated ancestors.

The field "is really blooming," agrees Zuzana Hofmanová, an EVA geneticist who is working on reconstructing the social structure of medieval Moravia using the bones of people buried under Czech and Slovak churches. "We're just scratching the surface. There's a lot to be done, and it's really fun."

AFTER THE FIRST COMPLETE ancient human genome was published in 2010, demonstrating the power of ancient DNA technology, a wave of studies mapped ancient migrations and population shifts. The work drew on samples widely scattered in time and space, because sequencing ancient genomes was at first so expensive that researchers often sampled just one or two individuals per site. Geneticists focused on broad patterns of similarity in the DNA, and archaeologists seeking fine-grained analyses of social structure and culture were sometimes sidelined. "We told stories from a biological perspective that were only indirectly related to actual history," says Joachim Burger, a geneticist at Johannes Gutenberg University Mainz. "Archaeologists always hated us for that, and they're absolutely right."

Today, sequencing can be done at a fraction of what it cost a decade ago, and often yields more information per sample. Researchers can analyze well-preserved DNA from an ear bone for as little as a few hundred dollars—about the same price as getting a radiocarbon date. Based on shared stretches of DNA, studies can identify siblings and link grandparents to grandchildren, tracing the chronology of a cemetery in a way traditional radiocarbon dating, accurate to several decades at most, cannot. "We often tell archaeologists to wait and do radiocarbon dating after we put together the pedigree," says University of the Basque Country geneticist Iñigo Olalde.

In the early days of ancient DNA, such kinship information was considered noise. As a Ph.D. student in 2015, Olalde recalls working out a way to identify first-degree relationships among ancient individuals—in order to weed siblings and parents out of his data sets. "If you want to understand how people move, you want as many unrelated individuals as possible. A family is just the same DNA all over again," Olalde says. "We weren't particularly interested in families."

He was surprised to find archaeologists felt differently. As ancient DNA technology improved, "we realized there was a lot of potential … it wasn't just father-son, or a couple of siblings," he says. "It was whole families, over multiple generations." Soon, he was partnering with archaeologists to spin samples originally gathered for sprawl-



A reconstruction of the 5700-year-old Hazleton North barrow in England, in which dozens of family members were buried over generations.

ing population studies into papers focused on prehistoric kinship.

IN THE LATE 1970S, dozens of skeletons were found under a 50-meter-long barrow called Hazleton North, built 5700 years ago by some of the earliest farmers in Great Britain—part of the same population that began to construct Stonehenge 700 years later. For decades, archaeologists wondered whether Hazleton and similar mounds were like family crypts or functioned more as shared shrines, collecting biologically unrelated members of the community.

DNA from the tomb has now provided clear answers: Hazleton was a family affair. Working with a team that included Newcastle University archaeologist Chris Fowler, Olalde showed the tomb contained 27 relatives spanning five generations, as well as seven unrelated people. Fifteen of the men in the tomb were related by descent to a single man, whereas none of the adult women were. Although the family tree included two young girls, adult women descended from the founder were conspicuously missing (see graphic, below).

That suggests that among these Neolithic Britons, women were buried with the family of their mates, not their parents—an echo of arrangements in the later El Argar culture that opens the possibility that the people at Hazleton were patrilocal, too. "In the absence of genetics, we just couldn't tell," Fowler says. Now, "We can combine biological sex with what we know about age to see that socially this was something significant." Archaeologists had speculated about patrilocality in past societies based on ethnographic evidence and clues from bone chemistry, but had never been able to conclusively demonstrate it. Fowler hopes studies of isotopes in the bones, which can serve as a geographic fingergprint, can show how those men and women moved during their lifetimes.

Even though the women at Hazleton were unrelated, they apparently helped structure the extended families. The family's founding father had four female partners. They were apparently from different generations, because a later woman in the tomb had children with the founder's son as well as his grandson by a different partner, suggesting the son and grandson were somewhat similar in age. So the team guesses the founder's four unions were sequential, spread over his lifetime.)

Over the course of a century or more, burials were organized according to the female line. Two women and their descendants were buried in the tomb's southern chamber, whereas offspring of the other women were buried in a separate passage on the other side of the mound. "When you look at the layout … it seems to really matter which of those first generation women you were related to, and it seems to have been followed for five generations," Fowler says. "That male lineage unifies the tomb, whereas female ancestors subdivide the community."

This summer, the longest and largest pedigree yet was published, from a Neolithic site in France called Gurgy 'les Noisats,' where nearly 100 individuals have been sequenced. The site predates Hazleton by 1000 years, and it provides strong evidence for patrilocality: Dozens of people over seven generations traced their descent back to one man. To date, most of the dozen or so prehistoric family trees from Neolithic and Bronze Age Europe suggest patrilocality. "It's surprising how consistent it's been," given the variety of social structures seen in recent societies, says Ian Armit, an archaeologist at the University of York who wasn't part of the Gurgy team.

Others caution that a handful of studies from a 5000-year period across Europe isn't enough to show patrilocality was a rule. "We're still looking at single-site studies that don't allow us to draw strong conclusions," says Wolfgang Haak, a geneticist at EVA.

Some DNA from Gurgy posed riddles. There were no half-siblings, strange at a time when people died early and partners could have taken another mate. The absence suggests an avoidance of even

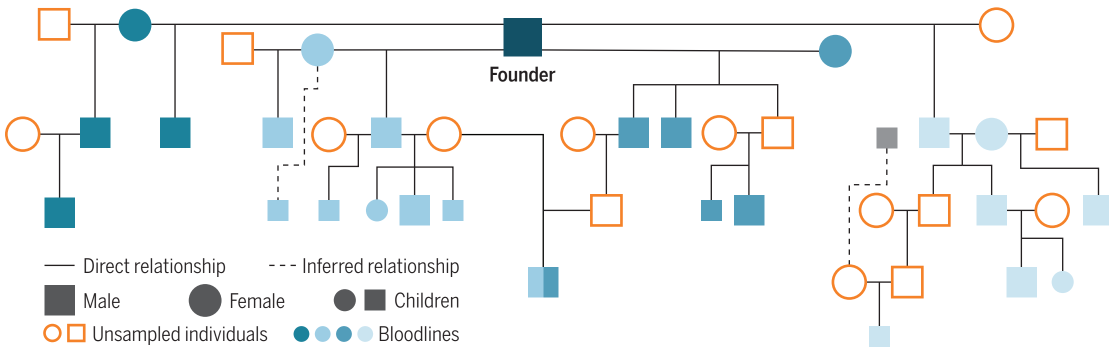
A Neolithic family affair

The complex family tree of a Neolithic clan buried in a chambered tomb called Hazleton North shows the promise of DNA to reveal ancient relationships. The founder was one man; shades of blue indicate his descendants with each of four female partners. The tomb holds many of his male descendants, along with their young children of both sexes (smaller symbols). Adult daughters are missing, presumably buried elsewhere. One third generation man descended from the founder on both sides.



Hazleton burial site

A 50-meter-long earthen mound was built as a tomb 5700 years ago near the modern town of Cheltenham, England.



serial partners, says Ghent University geneticist Maïté Rivollat. "Maybe they had rules we don't know about in terms of who was buried there, or maybe they were strictly monogamous."

At Gurgy, the team also did a new kind of analysis, adapting a technique commonly used by commercial ancestry testing companies to identify distant cousins. Called "identical by descent," or IBD, analysis, it compares the frequency and length of identical stretches of DNA to find more distant relatives. Standard ancient DNA analysis can connect people two degrees apart at most—a grandmother to her grandson, for example. IBD analysis allows geneticists to trace distant cousins, great-great-grandparents, or beyond.

IBD analysis showed how women influenced Gurgy's patrilocal community. "We see one woman leaves the community and a couple of generations later her descendants come back in," Ringbauer says. For example, a woman who had no daughter buried in the cemetery did have maternal descendants buried there. The missing female relatives must have left and borne offspring elsewhere; those offspring's descendants later returned and were buried at Gurgy. That suggests women stayed connected to their home village or birth family as part of a larger regional network.

Researchers are beginning to apply IBD to other sites and time periods. In a recent preprint, for example, Ringbauer identified second cousins who were buried 5000 years ago—and 1400 kilometers apart—in southern Russia and central Mongolia. Both men were members of the Afanasievo culture, an eastern variant of the steppe Yamnaya people who made a major genetic contribution to European populations around that time. At least one of the cousins must have traveled hundreds of kilometers across the steppe during his life, an early example of dramatic mobility in the space of a few generations.

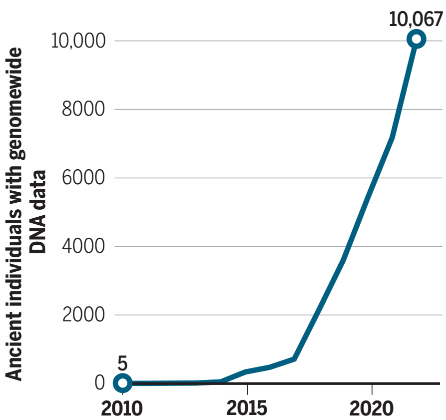
As ancient DNA databases grow exponentially, Ringbauer hopes access to tens of thousands of published genomes will create a sort of 23andMe for the distant past, linking prehistoric people to one another across time and space. "We can now screen tens of thousands of ancient individuals for IBD," Ringbauer says. "My vision is you'd add a study and immediately see, 'Oh, there's a second cousin over there.'"

PERHAPS CHASTENED by past criticisms, geneticists are now quick to concede that biological relatedness is only one element of kinship. Ethnographic studies of recent societies are rich in examples of biological parentage taking a back seat to other kinds of family relationships. They include matrilineal systems in which the most significant male relative is the mother's brother and communities where fatherhood is determined by the transfer of cattle to a woman's family. "Kinship is not just genetically determined, it's socially constructed," says Joanna Bruck, a University College Dublin archaeologist. "There has to be an acknowledgement that genetics can be read in multiple ways."

Archaeology can help guide those readings, as an early study of biological relatedness showed. It scrutinized the genomes of 84 people buried in small cemeteries up and down Germany's narrow, fertile Lech Valley between 2500 B.C.E. and 1650 B.C.E., during the technological transition from Neolithic farming to the Bronze Age. Published in 2017, the results showed patrilocality. But many Lech Valley cemeteries included both biologically related individuals and those with no genetic links to others in the graveyard.

A rising tide of ancient genomes

Thanks to better methods, as of November 2022 geneticists have genomewide data on more than 10,000 individuals. Researchers expect many more to come.



GRAPHIC: D. AN-PHAM/*SCIENCE*

Material culture helped explain the pattern. The unrelated people had many fewer grave goods, says Alissa Mittnik, a geneticist at EVA who co-authored the study. "The farmsteads were probably run by families who accumulated wealth. The others might have been farm workers, servants, or enslaved people." Those people did not establish families in one place across generations, but were buried in the plots of the presumably wealthier landowners, who did.

To puzzle out such community relationships, geneticists and archaeologists are forging stronger ties. For example, when Ludwig Maximilian University of Munich archaeologist Philipp Stockhammer, a lead author on the Lech Valley paper, was awarded a European Research Council grant to look for kinship patterns in Bronze Age Greece and the Aegean, his first step was to reach out to Greek archaeologists, asking for sites and questions where DNA could contribute. "Kinship analysis forces every lab to work closer with local archaeologists," Stockhammer says. "You can't do without archaeologists anymore."

Earlier this year, Stockhammer's team found that one-third of its samples from another Aegean project had long stretches of shared DNA, indicating they were the offspring of first cousins. That rate of close-kin interbreeding is highest documented anywhere in the past, although some modern societies have similarly high levels of cousin marriage.

Although the practice was slightly more frequent on isolated islands, it was seen across the Aegean, suggesting it was the result of a social practice or choice, rather than a lack of potential partners.

"Why do you have long-distance mobility of marriage partners in Central Europe in the Bronze Age, but not in Greece?" Stockhammer asks. Culture may hold an answer: The Bronze Age people of the Aegean grew olives and grapes. Those crops take decades to establish, offering an incentive to keep land in the family. "This kind of data lets us see how certain practices compare across regions," Stockhammer says. "The future of archaeogenetics is going to be zooming in, and then zooming out."

Paleogeneticists also hope to broaden their work beyond Europe, where research has flourished thanks to adequate funding, access to samples, and good preservation in the continent's relatively cool caves and burials. (See p. 53 for a review of research outside of Europe.) The work could also inform modern medicine, if researchers can identify a genetically inherited disease and trace it back generations. "Pedigree studies on modern individuals have been useful in tracing rare variants and how they're inherited," Hofmanová says. "Nine generation pedigrees from the past—and soon longer—will enable you to trace rare variants much further than we can in modern individuals."

In 2008, Haak was one of the first to apply ancient DNA to identify parents and children in a central German grave. Fifteen years later, the scope of information that DNA can reveal about life in the past is rapidly widening. "The question when you approach every grave is: Who are they? Are they related? Does proximity in space mean kinship?" Haak says. "Getting closer to answering some of those questions is a long-held dream that came true." ■



A fisherman in Ōkuma, Japan, fishes from a sea wall built after the tsunami in 2011 led to meltdown of the nearby Fukushima Daiichi nuclear plant.

PERSPECTIVES

ENVIRONMENT

The risks of radioactive wastewater release

The wastewater release from the Fukushima Daiichi nuclear plant is expected to have negligible effects on people and the ocean

By **Jim Smith[1], Nigel Marks[2], Tony Irwin[3]**

In 2011, the east coast of Japan suffered an earthquake and tsunami that resulted in the meltdown of three of the reactors at the Fukushima Daiichi nuclear plant. This led to an uncontrolled release of large amounts of radioactive material to the surrounding land and to the Pacific Ocean. More than 10 years later, new releases of radioactive wastewater to the Pacific Ocean from the Fukushima plant have started. The historical ocean contamination from Fukushima included large quantities of long-lived radioactive cesium-137 (^{137}Cs), so there is widespread concern over these new wastewater discharges. However, these releases will result in much lower total radiation doses to people and aquatic ecosystems than the contamination caused by the accident (*1*–*3*). Furthermore, aquatic ecosystems, including those around the Chernobyl disaster site, have been shown to be remarkably resilient to radiation (*4*–*7*). Thus, the new Fukushima releases are not expected to cause substantial effects on seafood consumers or the marine ecosystem.

The highest-activity radioactive contaminant in the Fukushima wastewater is tritium (^{3}H), in the form of tritiated water (HTO). This molecule cannot be separated from the wastewater because its chemical behavior is the same as that of nonradioactive water. Like other radionuclides [such as natural carbon-14 (^{14}C) and anthropogenic ^{137}Cs], which emit high-energy γ rays and β particles when they decay, tritium can have biological effects on organisms, particularly DNA damage (*8*). But tritium's radiotoxicity by ingestion is very low compared to that of these other radionuclides owing to its very weak β emission and relatively short retention time in the body. Therefore, it is common practice for nuclear facilities worldwide to discharge wastewater containing HTO into the sea (see the figure).

The La Hague nuclear facility in France annually discharges ~10,000 terabecquerels (1 TBq = 10^{12} Bq; the becquerel is an amount of a radioactive element that produces one decay per second) of HTO into the English Channel, with annual discharges during 1996–2016 ranging between 8000 and 12,000 TBq (*9*). Radiation doses to people (measured in sieverts and defined as the amount of energy deposited in tissue, taking account of the response of human biology to different radiation types) from the release of HTO from La Hague are low [<0.01 microsieverts per year (μSv/year) (*9*), which compares with the 1000 μSv/year recommended limit for members of the public from nuclear site releases (*2*)], and no environmental impacts have been found or are expected. For example, studies of fish (fathead minnow, *Pimephales promelas*) exposed to levels of HTO between 7700 and 21,900 Bq/liter in a river near a Canadian nuclear site (compared with <50 Bq/liter in seawater near La Hague) have indicated that it can induce DNA damage (as measured by comet assay), but no effects were observed on the survival of fish or various indices of health (*10*).

At Fukushima, 22 TBq of HTO per year will be released (*1*, *2*), which is a factor of ~450 less than that from La Hague per year. The planned Fukushima discharge annual limit is the same as when the facility was an operational electricity-generating boiling water reactor (BWR). Less tritium is produced in a BWR than in other types of nuclear reactor, so the limit was originally set to be low compared with that of some other reactor types. The very low release limit for the Fukushima wastewater was selected by the Japanese government owing to societal concerns such as the reputational impact on the local fishing industry. The planned releases will therefore take much longer, ~30 years, than strictly necessary according to current radiation safety recommendations.

There will be radionuclides other than tritium in the Fukushima discharge. The wastewater has been treated by passing it through a series of columns in which ion-exchange chemical reactions remove most of the radioactive elements that the wastewater con-

[1]University of Portsmouth, Portsmouth, UK. [2]Curtin University, Perth, Australia. [3]Australian National University, Canberra, Australia. Email: jim.smith@port.ac.uk

PHOTO: ONAS GRATZER/LIGHTROCKET

tains. It will be treated again before release if radiation levels are higher than the discharge limits. But tiny amounts of 30 other radionuclides remain after treatment, including ^{14}C, ^{60}Co, ^{90}Sr, ^{137}Cs, and Pu isotopes (*1*, *2*). In this regard, the Fukushima release is no different from routine releases from other nuclear sites around the world. For example, in 2019, the Sellafield nuclear site in the UK discharged 19 times more HTO, 1600 times more ^{14}C, 320 times more ^{60}Co, 180 times more ^{129}I, and over 1000 times more ^{90}Sr, ^{137}Cs, and Pu isotopes (*1*, *11*) than is planned annually for the Fukushima release. In 2019, increased radiation doses to seafood consumers near Sellafield were low and mainly due to historical discharges of natural radioactivity from a nearby non-nuclear site that produced phosphoric acid from uranium-rich phosphate ores, with the nuclear site contributing about a third of the dose of the phosphoric acid manufacturer (*11*).

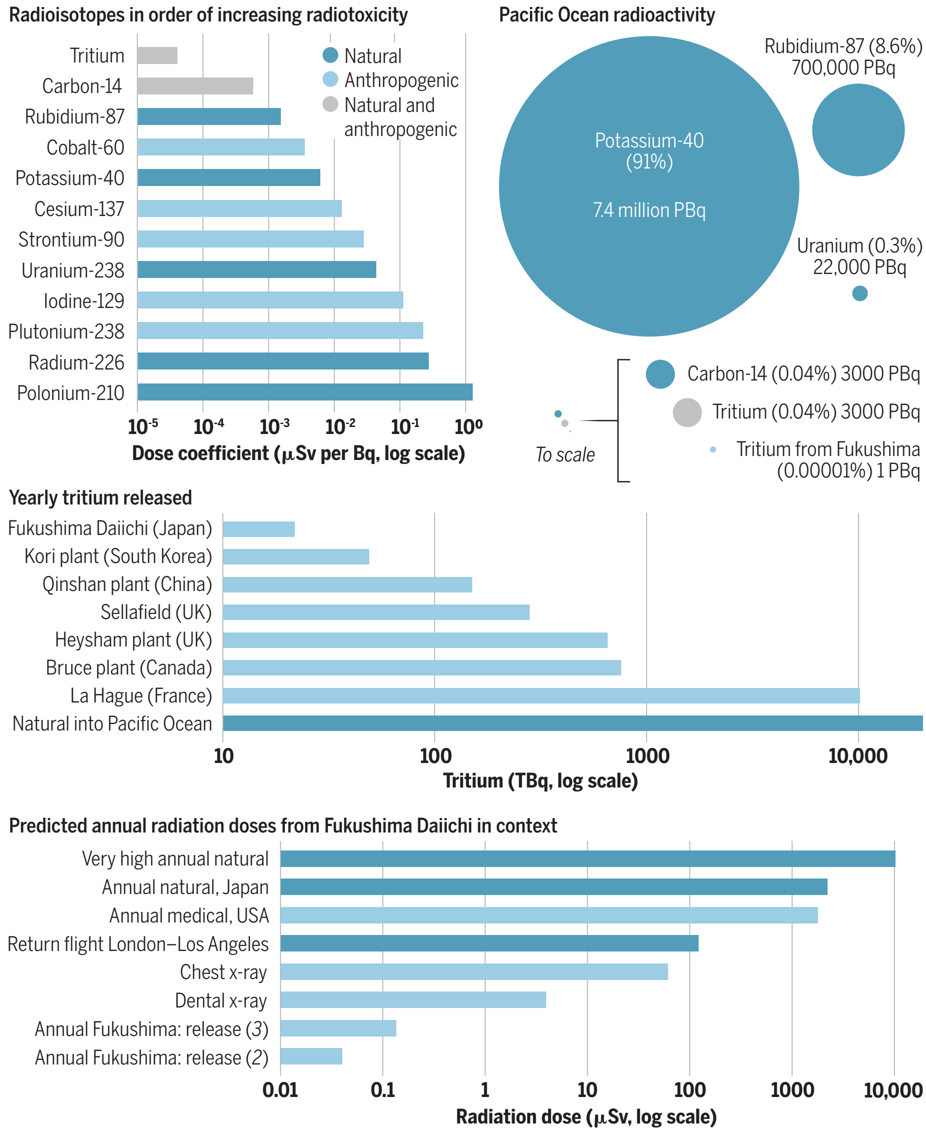
During discharge, the Fukushima wastewater is diluted 100 times so that tritium levels will be 2.5% of the Japanese Government regulatory limit and the sum of the 30 other relevant radionuclides will be less than 1% of the limits. The planned maximum of 1500 Bq/liter of tritium in the discharge water is ~14% of the 10,000 Bq/liter that the World Health Organisation designates as the drinking water limit for tritium (*12*).

Monitoring will take place to ensure that the levels of radiation in the Fukushima releases are below the regulatory limits (*1*, *2*). This includes analyses of the wastewater for all 30 radionuclides present, including tritium, total α and β radiation, and the energy spectrum of γ-emitting radionuclides. The process will be independently verified by the International Atomic Energy Agency (IAEA) by measuring wastewater, seawater, and aquatic organisms near the 1-km-long offshore discharge pipeline during the entire ~30-year discharge period (*2*). Since the first release began on 24 August 2023, publicly available information from Tokyo Electric Power Company (TEPCO), the IAEA and the Japanese Ministry for Agriculture, Fisheries and Food, reveals that tritium levels are negligible within 3 km of the site. Of 251 seawater measurements taken by 13 September at 10 sites within 3 km of the Fukushima plant, only one (10 Bq/l) is above the detection limit of ~8 Bq/l. No detectable tritium has been found in 25 fish at two sites 4 to 5 km North and South of the discharge point (limit of detection ~8 Bq/kg). Tritium is likely to be detectable by more sensitive methods and when releases are higher than that of the first batch release (~200 Bq/l of tritium).

Concerns have been expressed about bio-

Tritium from the Fukushima Daiichi releases in context

Radioisotopes, both naturally occurring and anthropogenic, exhibit different levels of radiotoxicity [based on (*15*)]. Tritium has a small dose coefficient mainly because its radiation is weak. In the Pacific Ocean, tritium contributes far less total radioactivity than naturally occurring radioisotopes. The planned annual release of tritium from Fukushima Daiichi is much lower than many releases from elsewhere, and the predicted annual radiation doses to local seafood consumers are much smaller than other sources of radiation.



Data are from the indicated references and publicly available information from the International Atomic Energy Agency, Woods Hole Oceanographic Institute, US Environmental Protection Agency, Japanese Ministry of Economy, Trade and Industry, the UN Scientific Committee on the Effects of Atomic Radiation, and the International Commission on Radiological Protection.

magnification—the increased concentration of pollutants in organisms—of tritium and its potential detrimental effects on marine life, but these concerns do not correctly reflect the risk. If, as planned, tritium is released in the form of HTO, it cannot biomagnify because its biological uptake and distribution follow that of the greater mass of water (*13*), which does not biomagnify. This lack of biomagnification is due to the chemical similarity of HTO and H_2O, which means that the ^{3}H/^{1}H ratio remains essentially unchanged in chemical and biological processes (*13*). Indeed, this is the very reason that HTO cannot be separated from ordinary water by Fukushima's water treatment facility. Tritium can be retained longer in organic molecules (and thus in organisms) as organically bound tritium (OBT) (*8*) when HTO is digested by aquatic organisms. But even assuming long exposure times, the radiation dose rates are vanishingly small.

Radiation doses of tritium and the 30 other radionuclides in the treated Fukushima wastewater have been calculated by independent researchers (*3*) and the IAEA

(*2*) using models that include dispersion in water, biomagnification, the different radiotoxicities, and longer retention times in organisms. These calculations result in very low dose rates (measured in energy deposited in tissue, Gy) to organisms from all radionuclides in the release: <0.005 µGy/hour (*3*) and <0.0001 µGy/hour (*2*). These levels are below the 40 µGy/hour dose rate that is considered to be the limit at which no ecosystem effects from damage to animal and plant tissues occur (*14*).

Aquatic ecosystems are surprisingly resilient to radioactive pollution because of the relatively high doses that are required to damage an organism's development and reproduction. Studies in lakes near Chernobyl and Fukushima found no effect of radiation on aquatic invertebrate abundance and diversity (*7*), or on developmental and genetic indicators (*4*, *5*). In perch (*Perca fluviatilis*), a slight delay in maturation of female gonads was associated with radiation levels in lakes at Chernobyl, but no effects were seen in roach (*Rutilus rutilus*), and fish condition indices were no different from those of uncontaminated lakes (*6*). Moreover, the fish population in the most contaminated lakes is diverse (*6*). Radiation dose rates in waterbodies at Chernobyl are more than 1000 times higher than those expected from the Fukushima discharges (*3*, *6*).

Radiation doses from the radioactive isotopes in the planned Fukushima releases to people, through all potential pathways including seafood consumption and bathing, are calculated to be less than 1 µSv/year (*2*, *3*). This is thousands of times smaller than the 2400 µSv/year global average radiation dose from natural background radiation. Risks from natural radiation primarily result from DNA damage by internal exposure to α and β radiation and to γ radiation originating inside or outside the body. So, the mode of action of naturally occurring radiation can be compared to that of the radioactivity from Fukushima. Natural radiation doses vary widely—millions of people worldwide receive doses >10,000 µSv/year. Current radiation protection models assume that any extra dose could potentially cause cancer, because all ionizing radiation could damage DNA. But a radiation dose of less than 1 µSv/year from the Fukushima release is not expected to have any public health consequence compared with, for example, natural radiation.

The impact of the Fukushima releases on the Pacific Ocean can also be understood by considering radioactivity in the natural environment. There is ~860 TBq of tritium in the Fukushima storage tanks, with 22 TBq discharged to the Pacific Ocean every year. The Pacific Ocean currently contains 500,000 TBq of natural and 2,500,000 TBq of anthropogenic tritium, the latter almost wholly from past atmospheric nuclear weapons testing. Furthermore, all tritium contributes just 0.04% to the total radioactivity of 8 billion TBq in the Pacific Ocean, where most radioactivity is due to naturally occurring potassium-40 (91%) and rubidium-87 (8.6%). It is important to note that there has long been awareness of radioactivity in seawater, but this has not been considered a health hazard.

Fears of radiation are likely to damage the livelihoods of Fukushima's fishing community, who are still recovering from fishing bans and reputational damage caused by the 2011 accident. To protect marine environments in the best possible way, resources and attention should be focused on key stressors including climate change, overfishing, and plastic pollution. The radiation protection science is clear that the Fukushima wastewater release presents no real threat to the organisms of the Pacific Ocean or to Fukushima's seafood consumers if carried out as planned. Any substantial deviation from the release plans would quickly be noticed by monitoring. Governments and researchers from elsewhere are also expected to closely monitor radioactivity in the Pacific Ocean during the release. The planned release at Fukushima is likely to be the most closely monitored wastewater discharge from a nuclear site. ■

REFERENCES AND NOTES

1. Tokyo Electric Power Company, *Radiological Impact Assessment Report* (2021); www.tepco.co.jp/en/hd/newsroom/press/archives/2021/pdf/211117e0102.pdf.
2. IAEA, *Comprehensive Report on the Safety Review of ALPS-Treated Water at the Fukushima Daiichi Nuclear Power Station* (IAEA, 2023).
3. R. Bezhenar, H. Takata, G. de With, V. Maderich, *Mar. Pollut. Bull.* **173** (Pt B), 112969 (2021).
4. N. Fuller *et al.*, *Environ. Pollut.* **292** (Pt B), 118479 (2022).
5. J. Goodman *et al.*, *Ecol. Evol.* **9**, 2640 (2019).
6. A. Lerebours *et al.*, *Environ. Sci. Technol.* **52**, 9442 (2018).
7. J. F. Murphy, L. L. Nagorskaya, J. T. Smith, *J. Environ. Radioact.* **102**, 688 (2011).
8. UNSCEAR, "United Nations Scientific Committee on the Effects of Atomic Radiation 2016 report: Report to the general assembly, with scientific annexes" (United Nations, 2017).
9. Institut de Surete et de Radioprotection Nucleaire, "Radiological report of the French environment from 2015 to 2017," IRSN DG/2018-00006 (2019).
10. B. Gagnaire *et al.*, *Sci. Total Environ.* **599-600**, 597 (2017).
11. Centre for Environment, Fisheries and Aquaculture, "Radioactivity in Food and the Environment, 2019 25th Edition" (2020).
12. World Health Organization, *Guidelines for Drinking-Water Quality* (WHO, 2022).
13. B. Fiévet *et al.*, *Environ. Sci. Technol.* **47**, 6696 (2013).
14. C. M. Larsson, *Ann. ICRP* **45** (suppl.), 41 (2016).
15. K. Eckerman, J. Harrison, H.-G. Menzel, C. H. Clement; ICRP, *Ann. ICRP* **41** (suppl. 1), 1 (2012).

ACKNOWLEDGMENTS

T.I. is the technical director of SMR Nuclear Technology Pty Ltd and chair of the Engineers Australia Sydney Nuclear Engineering Panel.

10.1126/science.adi5446

NEUROSCIENCE

Pregnancy programs the brain for mothering

Progesterone and estrogen exert separate effects on the brains of pregnant mice

By **Margaret M. McCarthy**

Pregnancy requires extensive physiological remodeling to provide the energy required for gestating a fetus and delivering and nurturing a newborn. Successful mothering requires an additional finely tuned behavioral repertoire. In most mammals, this starts with retrieval to a nest, grooming, and nursing. Estradiol and progesterone, the same pregnancy hormones that remodel the body, also program the brain for mothering, but precisely how this occurs has been opaque. On page 76 of this issue, Ammari *et al.* (*1*) report the identification of galanin neurons in the medial preoptic area (MPOA) as the cellular target of steroid programming in mice. Estradiol and progesterone changed the activity of galanin neurons such that the sensitivity to stimuli emanating from the pups (e.g., odors and vocalizations) was much stronger. New synapses were also formed on the galanin neurons. These remarkably specific changes were sufficient to promote mothering behavior.

Steroid hormones, such as estradiol and progesterone, are long-distance signaling molecules that are produced by endocrine organs (gonads, adrenals, and the placenta) and coordinate tissue responses across the body. Successful pregnancy and parturition require the orchestration of tissue remodeling of the mammary gland, uterus, and cervix along with construction of a placenta—all of which are responses to steroid hormones. However, the new mother also needs to respond to her newborns by providing protection, warmth, grooming, and nutrients through lactation. An absence of or a delay in this response can be catastrophic for species such as rats and mice, in which newborns have no capacity for thermoregulation or mobility. Synthesis of estradiol and progesterone is up-regulated during pregnancy, and

(2) using models that include dispersion in water, biomagnification, the different radiotoxicities, and longer retention times in organisms. These calculations result in very low dose rates (measured in energy deposited in tissue, Gy) to organisms from all radionuclides in the release: <0.005 µGy/hour (3) and <0.0001 µGy/hour (2). These levels are below the 40 µGy/hour dose rate that is considered to be the limit at which no ecosystem effects from damage to animal and plant tissues occur (14).

Aquatic ecosystems are surprisingly resilient to radioactive pollution because of the relatively high doses that are required to damage an organism's development and reproduction. Studies in lakes near Chernobyl and Fukushima found no effect of radiation on aquatic invertebrate abundance and diversity (7), or on developmental and genetic indicators (4, 5). In perch (*Perca fluviatilis*), a slight delay in maturation of female gonads was associated with radiation levels in lakes at Chernobyl, but no effects were seen in roach (*Rutilus rutilus*), and fish condition indices were no different from those of uncontaminated lakes (6). Moreover, the fish population in the most contaminated lakes is diverse (6). Radiation dose rates in waterbodies at Chernobyl are more than 1000 times higher than those expected from the Fukushima discharges (3, 6).

Radiation doses from the radioactive isotopes in the planned Fukushima releases to people, through all potential pathways including seafood consumption and bathing, are calculated to be less than 1 µSv/year (2, 3). This is thousands of times smaller than the 2400 µSv/year global average radiation dose from natural background radiation. Risks from natural radiation primarily result from DNA damage by internal exposure to α and β radiation and to γ radiation originating inside or outside the body. So, the mode of action of naturally occurring radiation can be compared to that of the radioactivity from Fukushima. Natural radiation doses vary widely—millions of people worldwide receive doses >10,000 µSv/year. Current radiation protection models assume that any extra dose could potentially cause cancer, because all ionizing radiation could damage DNA. But a radiation dose of less than 1 µSv/year from the Fukushima release is not expected to have any public health consequence compared with, for example, natural radiation.

The impact of the Fukushima releases on the Pacific Ocean can also be understood by considering radioactivity in the natural environment. There is ~860 TBq of tritium in the Fukushima storage tanks, with 22 TBq discharged to the Pacific Ocean every year. The Pacific Ocean currently contains 500,000 TBq of natural and 2,500,000 TBq of anthropogenic tritium, the latter almost wholly from past atmospheric nuclear weapons testing. Furthermore, all tritium contributes just 0.04% to the total radioactivity of 8 billion TBq in the Pacific Ocean, where most radioactivity is due to naturally occurring potassium-40 (91%) and rubidium-87 (8.6%). It is important to note that there has long been awareness of radioactivity in seawater, but this has not been considered a health hazard.

Fears of radiation are likely to damage the livelihoods of Fukushima's fishing community, who are still recovering from fishing bans and reputational damage caused by the 2011 accident. To protect marine environments in the best possible way, resources and attention should be focused on key stressors including climate change, overfishing, and plastic pollution. The radiation protection science is clear that the Fukushima wastewater release presents no real threat to the organisms of the Pacific Ocean or to Fukushima's seafood consumers if carried out as planned. Any substantial deviation from the release plans would quickly be noticed by monitoring. Governments and researchers from elsewhere are also expected to closely monitor radioactivity in the Pacific Ocean during the release. The planned release at Fukushima is likely to be the most closely monitored wastewater discharge from a nuclear site. ■

REFERENCES AND NOTES

1. Tokyo Electric Power Company, *Radiological Impact Assessment Report* (2021); www.tepco.co.jp/en/hd/newsroom/press/archives/2021/pdf/211117e0102.pdf.
2. IAEA, *Comprehensive Report on the Safety Review of ALPS-Treated Water at the Fukushima Daiichi Nuclear Power Station* (IAEA, 2023).
3. R. Bezhenar, H. Takata, G. de With, V. Maderich, *Mar. Pollut. Bull.* **173** (Pt B), 112969 (2021).
4. N. Fuller *et al.*, *Environ. Pollut.* **292** (Pt B), 118479 (2022).
5. J. Goodman *et al.*, *Ecol. Evol.* **9**, 2640 (2019).
6. A. Lerebours *et al.*, *Environ. Sci. Technol.* **52**, 9442 (2018).
7. J. F. Murphy, L. L. Nagorskaya, J. T. Smith, *J. Environ. Radioact.* **102**, 688 (2011).
8. UNSCEAR, "United Nations Scientific Committee on the Effects of Atomic Radiation 2016 report: Report to the general assembly, with scientific annexes" (United Nations, 2017).
9. Institut de Surete et de Radioprotection Nucleaire, "Radiological report of the French environment from 2015 to 2017," IRSN DG/2018-00006 (2019).
10. B. Gagnaire *et al.*, *Sci. Total Environ.* **599-600**, 597 (2017).
11. Centre for Environment, Fisheries and Aquaculture, "Radioactivity in Food and the Environment, 2019 25th Edition" (2020).
12. World Health Organization, *Guidelines for Drinking-Water Quality* (WHO, 2022).
13. B. Fiévet *et al.*, *Environ. Sci. Technol.* **47**, 6696 (2013).
14. C. M. Larsson, *Ann. ICRP* **45** (suppl.), 41 (2016).
15. K. Eckerman, J. Harrison, H.-G. Menzel, C. H. Clement; ICRP, *Ann. ICRP* **41** (suppl. 1), 1 (2012).

ACKNOWLEDGMENTS

T.I. is the technical director of SMR Nuclear Technology Pty Ltd and chair of the Engineers Australia Sydney Nuclear Engineering Panel.

10.1126/science.adi5446

NEUROSCIENCE

Pregnancy programs the brain for mothering

Progesterone and estrogen exert separate effects on the brains of pregnant mice

By **Margaret M. McCarthy**

Pregnancy requires extensive physiological remodeling to provide the energy required for gestating a fetus and delivering and nurturing a newborn. Successful mothering requires an additional finely tuned behavioral repertoire. In most mammals, this starts with retrieval to a nest, grooming, and nursing. Estradiol and progesterone, the same pregnancy hormones that remodel the body, also program the brain for mothering, but precisely how this occurs has been opaque. On page 76 of this issue, Ammari *et al.* (1) report the identification of galanin neurons in the medial preoptic area (MPOA) as the cellular target of steroid programming in mice. Estradiol and progesterone changed the activity of galanin neurons such that the sensitivity to stimuli emanating from the pups (e.g., odors and vocalizations) was much stronger. New synapses were also formed on the galanin neurons. These remarkably specific changes were sufficient to promote mothering behavior.

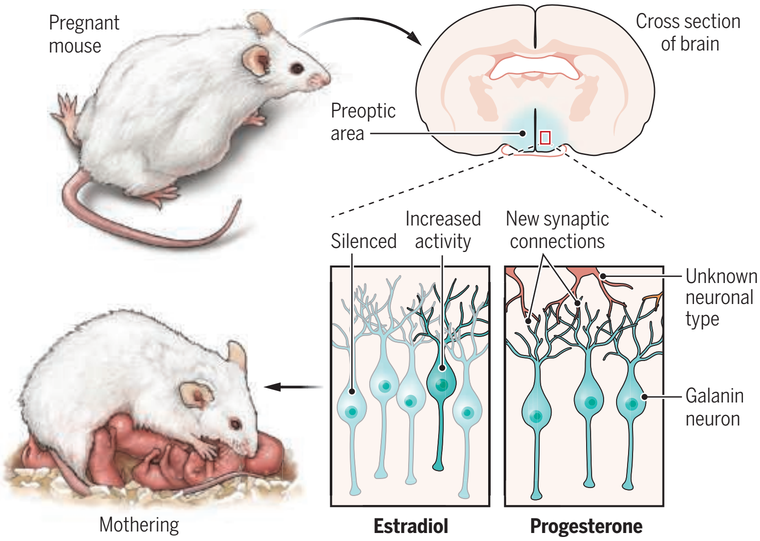
Steroid hormones, such as estradiol and progesterone, are long-distance signaling molecules that are produced by endocrine organs (gonads, adrenals, and the placenta) and coordinate tissue responses across the body. Successful pregnancy and parturition require the orchestration of tissue remodeling of the mammary gland, uterus, and cervix along with construction of a placenta—all of which are responses to steroid hormones. However, the new mother also needs to respond to her newborns by providing protection, warmth, grooming, and nutrients through lactation. An absence of or a delay in this response can be catastrophic for species such as rats and mice, in which newborns have no capacity for thermoregulation or mobility. Synthesis of estradiol and progesterone is up-regulated during pregnancy, and

these hormones act synergistically to promote reproductive behavior (mating and parenting). However, connecting steroid action to neuronal activity and behavioral modification has been challenging.

Steroids exert their effects, which can be long-lasting, through changes in gene transcription. To do this, estradiol binds to estrogen receptor 1 (ESR1), and progesterone binds to progesterone receptor (PR). ESR1 and PR are expressed at high concentrations in the MPOA, and the region plays an important role in controlling and coordinating sex differences in behavior, most notably mating and parenting (*2*). Neurons, astrocytes, and microglia of the MPOA exhibit different properties in males and females, including differences in morphology, synaptic density, and transcriptomic profile (*3*). Most neurons in the region are inhibitory—using γ-aminobutyric acid (GABA) as a neurotransmitter—and express ESR1. Galanin neurons are key regulators of maternal behavior (*4*), but they constitute only ~20% of ESR1-expressing neurons in the MPOA, and whether ESR1 and PR activation is required for the effects of galanin neurons on behavior was not known.

Hormones of pregnancy promote maternal behavior

Estradiol and progesterone levels increase during pregnancy and exert independent but coordinated effects on galanin neurons in the preoptic area of mice. Estradiol silences some galanin neurons and promotes the activity of others to increase the strength and selectivity of responses to stimuli emanating from newborn pups. Progesterone promotes the formation of new excitatory inputs, which endure past the time of immediate need for nurturing and thereby permanently change the behavior of female mice toward newborns.



Ammari *et al.* generated a mouse model in which *Esr1* and *Pgr* (which encodes PR) could be selectively ablated in MPOA galanin neurons. They did this in virgin mice and then allowed half of them to mate. The females who had pups showed a complete loss of normal nurturing behavior, with no retrieval to the nest or attempts at nursing. The authors determined that activation of ESR1 during pregnancy tuned the response of galanin neurons to pup stimuli by lowering the membrane potential and silencing most galanin neurons while selectively increasing the excitation of others, thus enhancing the signal-to-noise ratio. This fine-tuning faded as the pups matured. Activation of PR exerted a different effect by inducing excitatory synapse formation on the galanin neurons (see the figure). The identity of the neurons making new synapses onto the galanin neurons was not addressed. Unlike the transient nature of the ESR1-mediated effect, the PR-mediated changes endured and promoted nurturing behavior long after the original offspring were weaned. This observation may explain why females of many species behave maternally toward offspring that are not their own once they have experienced motherhood.

One of the important advances but also a remaining mystery from the work by Ammari *et al.* is that loss of *Esr1* or *Pgr* expression in the galanin neurons was sufficient to block the induction of maternal behavior by pregnancy. This was unexpected given the broad hormonal milieu of pregnancy, which includes increased oxytocin and prolactin, and because many other neuronal types express both steroid receptors. It may be that some component of pregnancy creates this specificity—a hypothesis that can be tested by inducing maternal behavior in virgins by other means (*5*). Regardless, the specificity is essential to avoid steroids promoting behaviors that are incompatible with parenting, such as mating or aggression (*6*). Future studies exploring the characteristics of neurons not responsive to the conditions of pregnancy will provide further information on the neural source of hormonally modulated behaviors. Moreover, deeper interrogation of the ESR1- and PR-responsive galanin neurons will help inform understanding of other aspects of mothering, including why it sometimes fails, resulting in females either ignoring or being aggressive toward pups. Equally important is how nutrients, stress, and substances of abuse affect the neural response to pregnancy and what epigenetic modifications might occur in these neurons.

Although mothering is arguably the fulcrum of reproductive success, a small percentage of mammalian species, including humans, benefit from fathering. Because fathers do not experience the hormonal milieu of pregnancy, it is not always clear how fathering behavior is induced. A study of more than 600 married men revealed a precipitous drop in circulating testosterone when they became fathers, and the more engaged that they were with their children, the lower the testosterone dropped, particularly during the first month after birth (*7*). By the time the child was 1 year old, testosterone levels were close to those of prefatherhood. Paternal behavior is much less studied than mothering in rodent models, but the findings of Ammari *et al.* provide information on where to begin, starting with determining if galanin neurons of the MPOA are activated during, and required for, fathering. If so, the next question would be whether testosterone fine-tunes the excitability of galanin neurons similar to estradiol and progesterone, which would suggest a common neural underpinning of parenting behavior. If the mechanism of fathering is different, it would be an example of compensation in which males overcome the limitation of not experiencing pregnancy and parturition through the creation of a separate neural process (*8*). The study by Ammari *et al.* also offers insights into how steroid signaling may be transduced to regulate competing behavioral repertoires, such as aggression, mating, and parenting by activating only a small subset of hormonally responsive neurons among the larger population. ■

Department of Pharmacology, Director UM-MIND, University of Maryland School of Medicine, Baltimore, MD, USA. Email: mmccarthy@som.umaryland.edu

REFERENCES AND NOTES

1. R. Ammari *et al.*, *Science* **382**, 76 (2023).
2. J. L. Goodson, *Horm. Behav.* **48**, 11 (2005).
3. M. M. McCarthy, *Annu. Rev. Neurosci.* **46**, 321 (2023).
4. Z. Wu, A. E. Autry, J. F. Bergan, M. Watabe-Uchida, C. G. Dulac, *Nature* **509**, 325 (2014).
5. D. S. Stolzenberg, F. A. Champagne, *Horm. Behav.* **77**, 204 (2016).
6. M. V. Wu, J. Tollkuhn, *Horm. Behav.* **95**, 3 (2017).
7. L. T. Gettler, T. W. McDade, A. B. Feranil, C. W. Kuzawa, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16194 (2011).
8. G. J. De Vries, *Endocrinology* **145**, 1063 (2004).

ACKNOWLEDGMENTS

M.M.M. receives financial support from National Institutes of Health grants R01MH52716 and R01DA039062.

10.1126/science.adk2495

CHEMISTRY

An oscillating reaction to produce clean fuels

Understanding chemical reaction mechanisms could help synthesize sustainable fuels

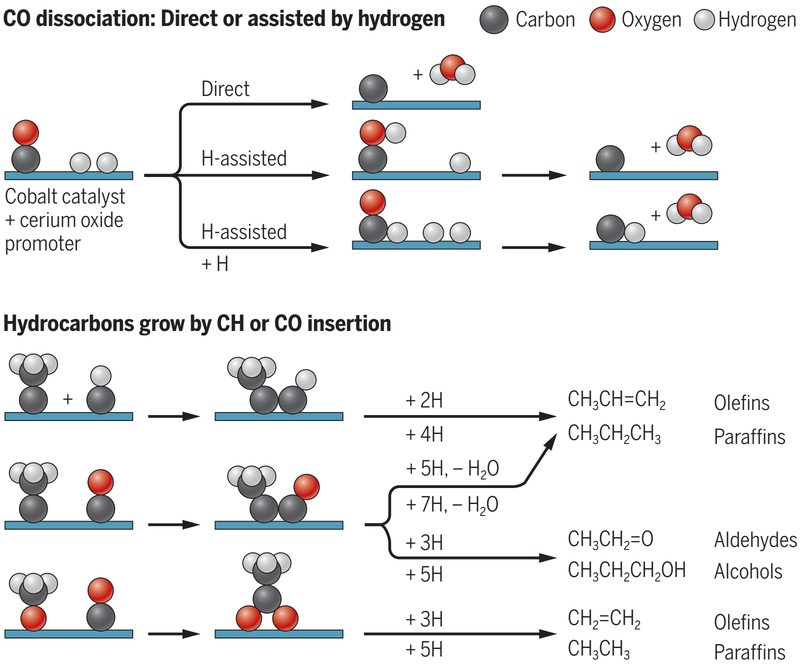
By **Hans Niemantsverdriet** *and* **Kees-Jan Weststrate**

Oscillating chemical reactions, in which reactions "flip-flop" reversibly, abound in living matter but are rare in the world of industrial chemistry. An oscillating reaction in solution, sometimes used for live demonstrations in chemistry classes, is the Belouzhov-Zhabotinsky reaction, in which color-changing patterns spread over a Petri dish (*1*). Oscillations may occur in chemical reactors that contain solid catalysts but are often elusive because they need to be synchronized over the entire volume to be observable (*2*). On page 99 of this issue, Zhang *et al.* (*3*) report previously unseen self-sustained oscillations of the Fischer-Tropsch synthesis (FTS), a complex reaction that produces hydrocarbon fuels from syngas [a mixture of carbon monoxide (CO) and hydrogen (H_2)]. The authors used a flow reactor packed with a cobalt catalyst and a cerium oxide promoter. Understanding why such oscillations arise may improve mechanistic insight of this process that features in future scenarios for developing sustainable fuels.

FTS is the reaction between CO and H_2 gases to form hydrocarbons, using cobalt or iron catalysts at moderate pressure (20 to 40 bar), and temperatures between 220° and 350°C (*2*). Discovered in 1925 and used in World War II to convert coal through syngas to hydrocarbon fuels for military applications, FTS is currently applied to convert natural gas and coal into ultraclean transportation fuels for cars and airplanes and into a versatile feedstock for the fabrication of more complex chemicals. FTS features prominently in future energy scenarios for storing electricity and for the production of sustainable aviation fuels, using captured CO_2 and "green" H_2 from water electrolysis with solar or wind power.

Elementary Fischer-Tropsch reactions on a cobalt catalyst

Carbon monoxide (CO) dissociates directly or is assisted by H atoms on the cobalt catalyst containing a cerium oxide promoter. Hydrocarbon chains grow by addition of methylidyne (CH) monomers or by insertion of CO followed by hydrogenation and desorption from the catalyst surface. The reaction rates and the hydrocarbon concentrations oscillate with time and according to temperature.



Mechanistically, FTS is a sequence of many elementary reactions on the surface of the catalyst, starting with the dissociation of CO and H_2, which form a monomer that polymerizes into paraffins, olefins, and oxygenated products. There are two main hypotheses on how the reaction proceeds (see the figure) (*4*). The first assumes that CO dissociates at atomic steps on the cobalt surface, after which C and H react to form hydrocarbons, and O atoms hydrogenate to water, which detaches from the catalyst. Water formation is the rate-determining step (the activation energy is 120 to 130 kJ/mol), whereas the direct splitting of the C–O bond requires considerably less energy (~100 kJ/mol) (*5*). Subsequent reactions to form hydrocarbons include the repeated coupling of CH monomers with alkylidyne (–CH–C_nH_{2n+1}) fragments, which is promoted by co-adsorbed CO, and proceeds with low activation barriers of 60 to 70 kJ/mol (*6*). Desorption of the hydrocarbon occurs at a higher activation energy of 80 to 100 kJ/mol, ensuring that hydrocarbon chain growth is favorable. The second hypothesis assumes that H atoms assist in breaking the C–O bond through a CHO(H) intermediate (*4*). Hydrocarbon chain growth is thought to occur from CO insertion into the growing chain, whereas O rejection happens later in the reaction sequence. Such CO insertion mechanisms account for the formation of, for example, alcohols and ketones, which form in small but detectable amounts. Computational modeling supports the feasibility of such routes (*4*), but experimental proof has not been reported so far.

Zhang *et al.* demonstrate that the intricate set of surface reactions constituting the FTS exhibit self-sustained oscillations in the concentration of reactants and products that go on for several hours, provided that the cobalt catalyst contains cerium oxide as a promoter (a material that facilitates a reaction without being a catalyst). Concentration oscillations occur between 200° and 260°C and 1 atm, implying that the hydrocarbon products are small and in the gas phase. The authors present a kinetic model in which CO is proposed to react with an alkoxy (R–O–) adsorbate to form a carboxylate key intermediate (R–COO) followed by partial hydrogenation to produce either an alkoxy intermediate for further growth or the final, oxygen-free hydrocarbon product. With a kinetic model that specifically includes the effect of temperature, they demonstrate that applying temperature variations as shown by the experiments, reproduces the observed oscillations in rate and product selectivity. However, this

GRAPHIC: K. HOLOSKI/*SCIENCE*

Syngaschem BV, Eindhoven, The Netherlands.
Email: jwn@syngaschem.com

model does not explain why such oscillations develop spontaneously, which requires further study.

Devising a kinetic model that reproduces the spontaneous oscillations observed by Zhang *et al.*—that is, without external perturbations such as temperature or pressure—is a challenge for which the literature offers inspiration (*7, 8*). For example, in addition to surface concentrations and rates of elementary reactions, factors such as exothermicity (released heat) and heat flow in the reactor need to be considered because these ultimately cause the oscillations in temperature and reactivity. Highlighting the sensitivity to heat transport, Zhang *et al.* report that oscillations disappear when they switch the diluting medium—present for analytical purposes—from argon to helium, which increases heat conductivity.

The role of cerium oxide, a material that binds and stores oxygen in automotive exhaust systems (*2*), is still unclear. In a scenario in which O removal is the rate-determining step, it could be envisaged that cerium suboxide [which contains cerium(III) (Ce^{3+})] attracts O atoms, accelerating the C–O dissociation and thus all subsequent pathways into hydrocarbons. Once the absorption capacity of cerium oxide is saturated, the reaction slows down, making the environment more reducing and causing reduction of Ce^{4+} back to Ce^{3+}, which then resumes its accelerating role until it is saturated again. Cleverly designed experiments to monitor the cerium oxidation state in situ may test this hypothesis. Likewise, experiments aimed at determining surface coverage of reacting species during the oscillations could also provide information on the reaction mechanism.

The finding of Zhang *et al.* that a catalytic system as complex as FTS may spontaneously develop oscillations that extend over an entire reactor is a notable result. It may open new avenues for exploratory research of oscillations in CO conversion and how they could offer a way to improve the overall performance of the FTS process, especially when synthesizing green fuels. ■

REFERENCES AND NOTES

1. B. P. Belousov, *Ref. Radiats. Med* **1**, 145 (1958).
2. I. Chorkendorff, J. W. Niemantsverdriet, *Concepts of Modern Catalysis and Kinetics* (Wiley-VCH, 2003)
3. R. Zhang, Y. Wang, P. Gaspard, N. Kruse, *Science* **382**, 99 (2023).
4. K. T. Rommens, M. Saeys, *Chem. Rev.* **123**, 5798 (2023).
5. C. J. Weststrate, P. van Helden, J. W. Niemantsverdriet, *Catal. Today* **275**, 100 (2016).
6. C. J. Weststrate *et al.*, *Nat. Commun.* **11**, 750 (2020).
7. R. Imbihl, G. Ertl, *Chem. Rev.* **95**, 697 (1995).
8. V. P. Zhdanov, *Surf. Sci. Rep.* **45**, 231 (2002).

ACKNOWLEDGMENTS

Syngaschem BV gratefully acknowledges discretionary funding from Synfuels China Technology Co. Ltd., Beijing-Huairou, P.R. China.

10.1126/science.adk5831



A site in White Sands National Park, New Mexico, USA, contains clearly identifiable ancient human footprints.

ANTHROPOLOGY

Dating the arrival of humans in the Americas

A debate about the age of ancient footprints continues

By **Bente Philippsen**

Dating the oldest evidence for the presence of *Homo sapiens* in the Americas is a matter of ongoing debate. One view is that the earliest such evidence is from 16,000 to 14,000 years ago, after the Last Glacial Maximum (LGM), when people would have crossed the Beringian strait from Siberia over a dry land bridge (*1*). However, controversial radiocarbon dating of seeds from a set of preserved footprints in the White Sands National Park in New Mexico, USA, indicates that humans were present in the Americas as early as 21,000 to 23,000 years ago—during the LGM (*2*). This site has already inspired a vivid discussion about dates and human migrations (*2–6*). On page 73 of this issue, Pigati *et al.* (*7*) present a new approach and independent data from the White Sands site that may be the most convincing evidence for pre-LGM presence of *H. sapiens* in America.

Evidence for the post-LGM migration of *H. sapiens* from Siberia to America is present at many archaeological sites and has left genetic traces in the current population of the Americas (*1*). However, this does not exclude the possibility of earlier migrations. Evidence for the peopling of other continents shows that *H. sapiens* were able to cross large distances over land and even the sea tens of thousands of years ago. The White Sands site contains several sediment layers with clearly identifiable human footprints (*2*). The study that produced the first estimate of 21,000 to 23,000 years for the age of the footprints used radiocarbon dating of the seeds of the aquatic plant ditchgrass (*Ruppia cirrhosa*) (*2*), which were embedded between sediment layers that contained the footprints. If this estimate is correct, the site is much earlier than other well-dated sites with clear evidence of human occupation and could be a convincing sign of human presence in America during the LGM (*1*).

Discussions about the White Sands site have mainly focused on the challenges involved in dating it. Radiocarbon dating is a well-established and widely applied method of age determination in archaeology. However, there are some potential sources of error that have to be taken into account. One such source is the freshwater reservoir effect (*3–6, 8, 9*). This occurs when submerged plants photosynthesize dissolved inorganic carbon from the water. The dissolved inorganic carbon can originate from atmospheric carbon dioxide and thus have the same age as that of contemporary terrestrial vegetation. It can, however, also originate from ancient carbon sources: groundwater that had been without contact to the atmosphere for millennia, dissolved carbonate minerals, or mineralized organic matter from peat bogs. Therefore, aquatic

National Laboratory for Age Determination, University Museum, Norwegian University of Science and Technology, Trondheim, Norway. Email: bente.philippsen@ntnu.no

plants such as *R. cirrhosa* can incorporate ancient carbon and thus appear older than they are when they are radiocarbon dated.

The difference between the plant's radiocarbon age and the radiocarbon age of contemporary terrestrial plants is called the reservoir age. In most cases, the dissolved inorganic carbon in water bodies has a multitude of sources that can lead to potentially very high and highly variable reservoir ages. The concentrations of the ^{13}C isotope in the dated samples should always be included because they can help to identify the carbon source(s) and would have been a valuable addition to the debate about the age of the footprints (*2*). Another potential source of error in (*2*) is the redeposition of the *R. cirrhosa* seeds. They could have been preserved in much earlier sediments and redeposited millennia later to form the seed layers in which the footprints were set.



Conifer pollen (average size >70 μm) were used for dating.

PHOTO: FLOW CYTOMETRY RESOURCE FACILITY, INDIANA UNIVERSITY

To avoid the freshwater reservoir effect, Pigati *et al.* performed radiocarbon dating of nonaquatic pollen grains from the White Sands site. With the aim of avoiding error introduced through redeposition, the authors also used optically stimulated luminescence (OSL) dating of the sediment itself [a solution to the debate that was suggested in (*4*)], which determines when mineral grains such as quartz were last exposed to light. The approach can therefore date when sediments were redeposited by wind and similar processes. Both methods confirm the age of the White Sands footprints as 21,000 to 23,000 years ago.

The approaches used by Pigati *et al.* are not without limitations. Radiocarbon dating of pollen grains is an intricate process that comes with a risk of contamination. OSL dates can have large measurement uncertainties, and factors such as background radioactivity and water saturation have to be known to calculate accurate ages. Furthermore, the quartz grains could have been shaded by other material such as gypsum, which is prevalent in the sediments of the White Sands site, and thus not be fully exposed to light during redeposition ("incomplete bleaching"). However, it is highly unlikely that all the error sources present across the Pigati *et al.* study and in (*2*) introduced the same magnitude of error. Thus, even allowing for substantial errors in radiocarbon dating and OSL, the data overall from the Pigati *et al.* study strongly indicate human presence in the Americas around the LGM.

The controversy surrounding the dating of the White Sands footprints provides some valuable insights for dating this and other potentially very old sites. In particular, the findings of Pigati *et al.* emphasize that the context of a sample is one of the most important pieces of information in radiocarbon dating. To optimize dating accuracy, the carbon source analyzed in a radiocarbon study should be as physically and temporally close to the evidence of human action as possible. In terms of dating precision, much can be gained from including stratigraphy, elevation, or dates from other methods in the radiocarbon age model. Such a combination of methods and materials can target several different mechanisms of carbon uptake and sample deposition. Although there is no guarantee for accurate dates even when using all of these strategies, it is highly unlikely that different methods applied to different materials will always result in systematic errors of the same direction and same order of magnitude.

Absolute dates for archaological sites are necessary to connect human migrations and cultural development with climate and environment. Future projects could include following the White Sands footprints to a potential campsite, which might yield further evidence of human occupation such as stone tools [suggested by (*3*)]. Furthermore, targeted sampling of additional materials for radiocarbon and other dating techniques, and DNA analysis of ancient bones and sediments on the White Sands site and other potentially early sites, could provide useful information. Results and materials from rescue and contract archaeology (such as archaeological investigation before road construction) should also be reevaluated and dated with the same scrutiny as those from research excavations. Last, the conclusions from the conversation around the dating of the White Sands footprints can help to inform future requirements for rescue and contract archaeology, such as deeper digs to access earlier layers, or the collection of dating samples for future research projects. ■

REFERENCES AND NOTES

1. M. R. Waters, *Science* **365**, eaat5447 (2019).
2. M. R. Bennett *et al.*, *Science* **373**, 1528 (2021).
3. C. V. Haynes Jr., *PaleoAmerica* **8**, 95 (2021).
4. D. B. Madsen *et al.*, *Science* **375**, eabm4678 (2022).
5. C. G. Oviatt *et al.*, *Quat. Res.* **111**, 138 (2022).
6. J. S. Pigati *et al.*, *PaleoAmerica* **8**, 99 (2022).
7. J. S. Pigati *et al.*, *Science* **382**, 73 (2022).
8. D. M. Rachal *et al.*, *Quat. Sci. Adv.* **3**, 100024 (2021).
9. D. M. Rachal *et al.*, *Geoarchaeology* **37**, 923 (2022).

10.1126/science.adk3075

MATERIALS SCIENCE

Damaging diamond with shock waves

Deformations inflicted in diamond could help design stronger materials

By **Marcus D. Knudson** *and* **Christopher T. Seagle**

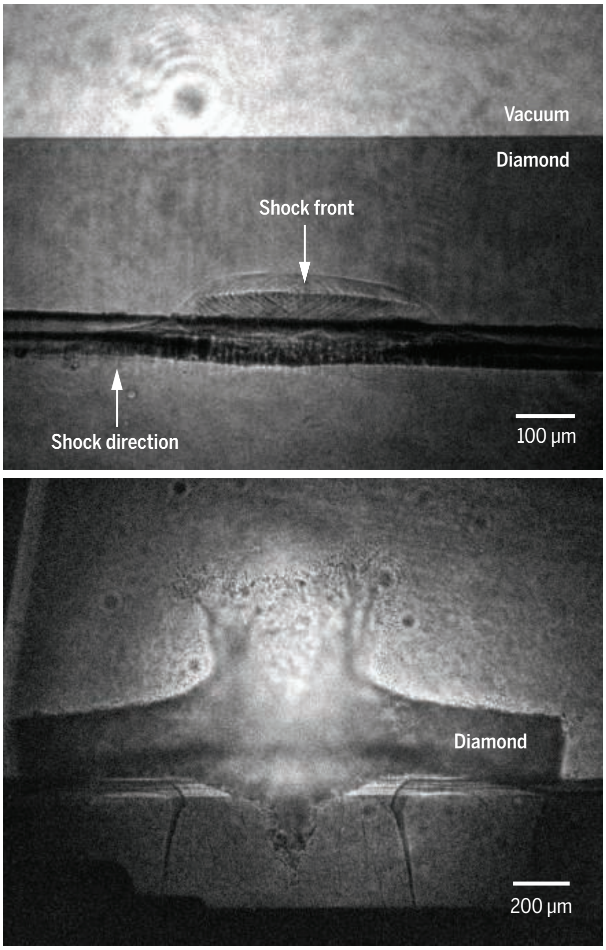
The most extreme thermodynamic conditions that can be produced in the laboratory are created by using shock waves. Shock waves are the waves generated in a material after the infliction of a mechanical impact or rapid deposition of energy, such as from a high-intensity laser. When a strong shock propagates within a single crystal of a material, it deforms the crystalline lattice. This deformation is accommodated by the creation and propagation of dislocations—crystal defects that contain an irregular arrangement of atoms within the lattice. On page 69 of this issue, Katagiri *et al.* (*1*) report imaging experiments on shocked diamond single crystals. Using an x-ray free-electron laser (XFEL) as an ultrafast probe, they observed shock-induced dislocation motion that was faster than the speed of sound in diamond, called transonic dislocations. This experimental evidence is an important benchmark for the theory of dislocation dynamics and could aid designing stronger materials.

Deformation induced by a shockwave is macroscopically restricted to the direction of shock propagation. At low stresses, below the material's elastic limit, this deformation is reversible and is also microscopically uniaxial, meaning, with no deformation in lateral directions. Shock amplitudes above the elastic limit generate sufficiently high stress that the material can no longer resist lateral deformation and undergoes irreversible plastic deformation. Details of dislocation dynamics—that is, the generation, interactions, and movement of crystal defects—provide information on the ability of a material to resist deformation (the material's strength) and its damage and failure response (how a material breaks) (*2*). Theory and simulation both provide insight

Sandia National Laboratories, Albuquerque, NM, USA.
Email: mdknuds@sandia.gov

into dislocation dynamics. Recent advances in high-performance computing have enabled simulations to investigate dislocation dynamics, including direct shock wave propagation (*3*). However, there has been a dearth of experimental studies testing the predictions of theory and simulation, particularly in real time, as shock waves propagate in a crystalline (or solid) material.

Deformation processes in crystalline materials that occur during dynamic compression are traditionally studied by using wave profile measurements or recovery experiments (*4*). Wave profile measurements use Doppler-frequency–shifted laser light reflected from a sample surface to infer its surface velocity—similarly to a radar gun measuring the velocity of a car on the highway. Wave profile measurements provide an integrated, macroscopically and spatially averaged response and are typically not sensitive to microscopic phenomena (including dislocation formation and propagation) associated with deformation mechanisms in solids. Recovery experiments, which allow the "interrogation" of shocked samples after the dynamic event using numerous microscopic techniques, can provide insight into deformation mechanisms. However, it is difficult to separate the dislocation dynamics occurring during the submicrosecond shock event with subsequent deformation on the seconds timescale, which is associated with shock recovery.



X-ray radiography images show a shockwave propagating in a diamond crystal at early (8 ns, top) and late (500 ns, bottom) times.

Recent advances in the coupling of dynamic compression drivers (controlled shock wave generators such as gas guns and large lasers) to high-brightness x-ray light sources have opened new windows into deformation studies during high strain-rate compression, allowing imaging of materials in time-resolved experiments. Over the past several years, dynamic compression capabilities have been added at synchrotron light sources such as the Advanced Photon Source (APS) (*5*) and the European Synchrotron Radiation Facility (ESRF) (*6*); and XFEL sources such as the Spring-8 Angstrom Compact Free Electron Laser (SACLA) (*7*) and the Linac Coherent Light Source (LCLS) (*8*). The extreme brightness and coherence of the light from these sources provide copious photons in subnanosecond bursts or pulses that are required to penetrate micrometer- to millimeter-scale samples during dynamic events.

The work by Katagiri *et al.* is an example of the types of experiments that have been enabled through these recent advances. The authors used an intense nanosecond-scale, visible-light laser to drive ~100-GPa shocks into single-crystal diamond samples, synchronizing the visible laser and the XFEL such that the shock occurred a few to tens of nanoseconds before the x-ray (XFEL) burst. Radiography measurements revealed substantial photon scattering off stacking faults (a type of planar defect) along specific crystallographic planes within the plastically deformed region of the sample. This enhanced scattering allowed them to "see" these planar defects within the crystal. Once initiated, these defects propagated, closely tracking the leading shock front. Combining multiple image frames enabled them to measure the dislocation motion (see the images). The inferred velocities reached the speed of sound and thus provide experimental evidence of transonic dislocation propagation. This is an important benchmark because it allows better calibration with theory and simulation results.

There are several reasons why Katagiri *et al*. focused their study on diamond. Diamond is a prototypical brittle ceramic and has long been of scientific and technological interest because of its exceptional mechanical, optical, and thermal properties. Critically for the experiments of Katagiri *et al.*, however, diamond also exhibits high transmission of low energy, so-called soft x-rays (with energies of less than ~10 keV). Although XFELs produce extremely high x-ray flux, the usable energy range is currently limited to a few to 10 keV. This restricts the types of materials that can be studied with these techniques.

This restriction should be lifted in the coming years. Major investments in high-brightness light sources, such as the LCLS (*8*) and APS upgrades (*9*), promise higher x-ray flux, and more notably, high flux at high energy, so-called hard x-rays (10 to 40 keV). This will enable similar types of investigations on higher-density, polycrystalline materials, in which dislocation dynamics on grain boundaries (that occur between micro- or nanocrystals) play a major role in determining material strength and resistance to damage (*10*). Enhanced scientific understanding of deformation mechanisms in crystals gained through these studies may aid in the design of engineered materials that are more resistant to deformation and failure. ■

REFERENCES AND NOTES

1. K. Katagiri *et al*., *Science* **382**, 69 (2023).
2. B. L. Hansen, I. J. Beyerlein, C. A. Bronkhorst, E. K. Cerreta, D. Dennis-Koller, *Int. J. Plast*. **44**, 129 (2013).
3. K. Nguyen-Cong *et al*., In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (2021), pp. 1–12.
4. G. T. Gray III, *Annu. Rev. Mater. Res*. **42**, 285 (2012).
5. X. Wang *et al*., *Rev. Sci. Instrum*. **90**, 053901 (2019).
6. N. Sévelin-Radiguet *et al*., *J. Synchrotron Radiat*. **29**, 167 (2022).
7. Y. Inubushi *et al*., *Appl. Sci*. **10**, 2224 (2020).
8. A. Halavanau, F.-J. Decker, C. Emma, J. Sheppard, C. Pellegrini, *J. Synchrotron Radiat*. **26**, 635 (2019).
9. Advanced Photon Source Upgrade Project Final Design Report, ANL Report APS-2.01-RPT-003.
10. C. A. Bronkhorst *et al*., *Int. J. Plast*. **118**, 70 (2019).

ACKNOWLEDGMENTS

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the US Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. Any subjective views or opinions that might be expressed do not necessarily represent the views of the US Department of Energy or the United States Government.

10.1126/science.adk4420

CELL BIOLOGY

Y chromosome proteins in female tissues

Unreliable protein-based tools are impeding sex-based research

By **Bradley D. Gelfand** *and* **Jayakrishna Ambati**

The role of sex in health and disease is of central concern to biomedical research, as reflected by policies from the US National Institutes of Health, European Commission, and Canadian Institutes of Health Research that require investigators to consider sex as a biological variable (*1*). Mechanistic understanding of the role of sex in human health is essential to develop better and more-personalized care (*2*, *3*). Because the inclusion of both sexes in basic, translational, and clinical research is now the rule rather than the exception, it is necessary to examine the mechanisms responsible when sex differences are inevitably observed. Protein-based resources, which make up an essential mechanistic research toolset, have not been developed for the specific challenges presented by sex chromosome–encoded proteins. Consequently, an alarming number of protein-based resources are erroneous, inaccurate, or misleading by, for example, reporting Y chromosome–encoded protein expression in cells and tissues that lack a Y chromosome.

The contemporary theory of mammalian sexual differentiation (*4*) establishes a framework for identifying mechanistic causes of sexual dimorphism (phenotypes that occur more often in a particular sex). This theory states that sexually dimorphic phenotypes can largely be attributed to the three basic categories of sex-biasing factors: permanent or organizational effects of gonadal hormones, which typically happen around the time of birth or before (e.g., testosterone causes formation of a penis and scrotum in fetuses); reversible or activational effects of gonadal hormones in adulthood, or at the time of the experiment (e.g., the transient anti-inflammatory activities of estrogens); and sex chromosome effects owing to the unequal cell-autonomous activities of X and Y gene expression in, for example, XX and XY cells.

The influence of sex hormones on biological processes cannot be overstated, but strong evidence points to sex chromosomes playing crucial roles in human health beyond reproductive organization and function. For example, Jacobs syndrome, caused by an extra copy of chromosome Y (XYY), is associated with increased prevalence of asthma, autism spectrum disorder, and seizures (*5*). The extent to which disorders of the complement of sex chromosomes exhibit presentations beyond reproductive function indicates that sex chromosomes can contribute much more broadly to human health. Recent work on the somatic mosaic loss of the Y chromosome further indicates key roles for Y chromosome–encoded genes in a variety of complex diseases and in human longevity (*6*, *7*), including in cardiac fibrosis and mortality (*8*). Experimentally elegant genetic tools have revealed key roles for sex chromosomes in a host of biologic and pathologic processes, including behavior, cardiovascular diseases, immunity, metabolism, and Alzheimer's disease (*9*). In addition to chromosome-wide disorders, widespread transcription of individual Y chromosome–encoded genes outside of reproductive organs has been reported, and evidence indicates that individual gene dosage differences may contribute to underlying sex differences (*10*).

These observations highlight that sex chromosome–encoded genes may contribute to well-documented dimorphism in the incidence, severity, and even treatment efficacies for human diseases. Identifying the specific gene or genes on sex chromosomes that are responsible may unlock new therapeutic paradigms to restore, enable, or silence sex-specific gene function. For example, it may be possible to reduce the incidence of disease in females to match that of males by introducing a cohort of Y chromosome–encoded genes into affected cells and tissues. Unlocking this potential requires a detailed mechanistic understanding of sex chromosomes in human health through molecular characterization of gene expression and protein function.

Investigating gene function for sex chromosome–encoded genes presents distinct challenges. A cohort of sex chromosome genes includes ancestrally related X and Y chromosome gene pairs, or gametologs, which reside on the X and Y chromosomes. These gene pairs encode proteins with amino acid sequence homologies that can exceed 90% (*10*), indicating that, at least in part, they have overlapping functions (see the figure). However, emerging evidence suggests that these X-Y gene pairs can exhibit distinct activities that may contribute to sexually dimorphic phenotypes. For example, despite 92% amino acid identity with its X chromosome gametolog, DEAD box helicase 3 X-linked (DDX3X), the Y chromosome–encoded DDX3Y exhibits weaker adenosine triphosphatase (ATPase) activity, greater propensity for liquid-liquid phase separation, and altered RNA metabolism (*11*). The extent to which differences in DDX3 or other highly related gametologs contribute to sexually dimorphic phenotypes is not yet known.

Many genomic and transcriptomic resources accurately distinguish human gametologs. For example, in the Genotype-Tissue Expression Project, a comprehensive atlas of human mRNA expression, tissues from female donors show almost no transcripts mapping to Y chromosome–encoded genes (www.gtexportal.org). Nonetheless, it is important to be vigilant that mapping of transcripts to sex chromosomes is accurate and does not result in over- or underestimates of the expression of either gametolog.

Unfortunately, many protein-based resources are unable to distinguish between X and Y chromosome–encoded gametologs. A survey of commercial antibodies purporting to recognize Y chromosome–encoded proteins identified more than 60 such antibodies with supplier-provided marketing materials that report positive immunoreactivity in female cells and tissues (*12*). It is also troubling that the Human Protein Atlas, an open-access program to map all human proteins in tissues, reports that female tissues display expression of Y chromosome–encoded genes, including, for example, gliomas from females with positive immunostaining of neuroligin 4 Y-linked (NLGN4Y); breast cancers from females positive for zinc finger protein Y-linked (ZFY); and ubiquitously transcribed tetratricopeptide repeat containing Y-linked (UTY) expression in the neurons of normal female brains (*13*).

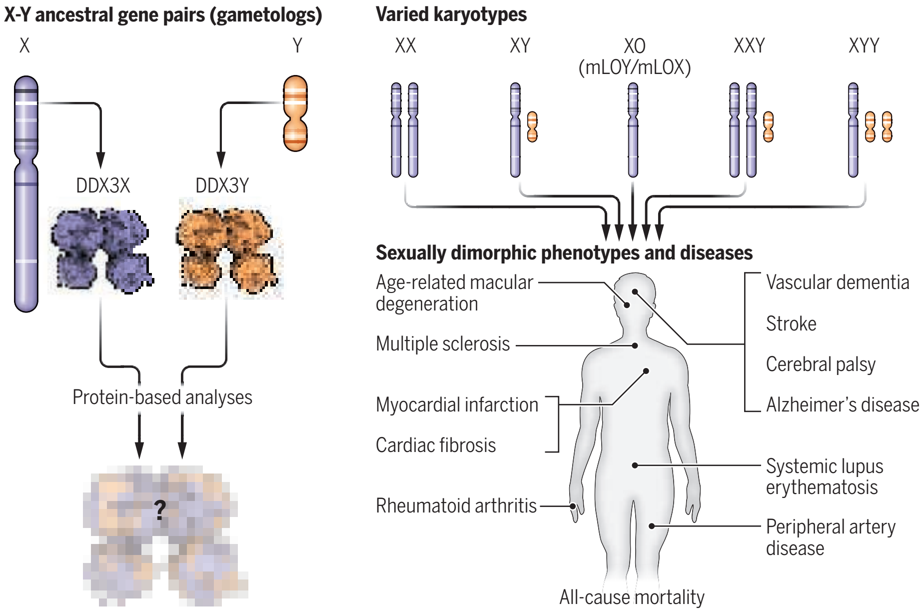
Some female tissues may have bona fide Y chromosome–derived antigens in cases of microchimerism, in which an individual harbors a small number of cells from genetically distinct individuals by way of a preg-

Center for Advanced Vision Science, University of Virginia School of Medicine, Charlottesville, VA, USA. Email: gelfand@virginia.edu

nancy with an XY fetus, a dizygotic twin, organ transplant, or blood transfusion (*14*). However, according to the survey (*12*), most of these primary antibodies are validated in female monoclonal cell lines, and data from the Human Protein Atlas demonstrate Y chromosome protein immunoreactivity in neurons of normal brain tissue (unlikely to have been the subject of transplantation). The most parsimonious explanation is that these represent false positive signals arising from poorly characterized antibodies.

The role of sex chromosomes in disease

Although gametologs encoded on the X and Y chromosomes are highly homologous, they may have distinct functions [such as DEAD box helicase 3 X-linked (DDX3X) and DEAD box helicase 3 Y-linked (DDX3Y)]. Typical protein-based assays are unable to distinguish between gametologs, thwarting research of sexual dimorphism in disease. The copy number of gametologs in XX, XY, XO (Turner syndrome), mosaic loss of Y chromosome (mLOY), mosaic loss of X chromosome (mLOX), XYY (Jacobs syndrome), and XXY (Klinefelter syndrome) may contribute to a variety of phenotypes or diseases whose prevalence, severity, symptomatology, or risk factors are sexually dimorphic.



This problem also extends beyond immunoassays because there are many publicly deposited protein sequencing studies with peptides mapped to Y chromosome–encoded genes in female cells and tissues. For example, proteomic analysis of the Cancer Cell Line Encyclopedia reports expression of Y chromosome–encoded ribosomal protein S4 Y-linked 1 (RPS4Y), lysine demethylase 5D (KDM5D), DDX3Y, UTY, and ubiquitin-specific peptidase 9 Y-linked (USP9Y) in female cancer cell lines (*15*).

Reliable and accurate protein resources are essential for understanding expression and function of all genes, and especially sex chromosome–specific genes. These problems could be addressed in several ways. It is important to more precisely categorize and label sex-specific peptides and proteins in publicly available antibody labeling and mass spectrometry data. This improvement can be implemented without developing new technologies but by better annotating existing protein expression databases and antibody datasheets for X and Y chromosome–encoded genes. Antibodies targeting gametologs should be assumed to be incapable of distinguishing between X and Y chromosome gene pairs unless explicitly validated for such applications. Antibody suppliers should relay this information, and researchers should state this explicitly when discussing results. With respect to other resources of protein expression data, caveats about the relative confidence and ambiguity of peptide mapping between X-Y gene pairs should be made explicit. This will improve the accuracy and transparency of existing materials, methods, and interpretations thereof.

In addition to research specifically targeted toward understanding the biology of sex chromosome–based differences, gametologs should be used to stress test the limits of specificity and sensitivity for protein identification and quantification more broadly. The existence of Y chromosome–derived proteins in female cells in the Cancer Cell Line Encyclopedia (*15*) implies that other protein pairs or families with similarly high homology are probably being misidentified. Although such misidentifications would be more pernicious to detect, the proportion of misidentified gametologs could serve as a proxy for the rate at which other high-homology pairs are being misidentified. Using this benchmark false positive rate could be harnessed to classify the likelihood of mismatched protein identification for every target in every sample type. The acknowledgment of such specificity limitations by producers and consumers is key to the accurate and transparent use of these resources.

Validated resources should be created that can precisely distinguish between sex chromosome–specific proteins, including antibodies and protein mass spectrometry protein identification and quantification algorithms. This may require the development and implementation of new approaches for validating antibody and mass spectrometry throughputs. For example, antibody production should design antigenic peptides toward regions of sequence dissimilarity among gametologs, cross-adsorb antisera against the opposing gene pair member, and, in the case of Y chromosome–encoded antigens, validate negative antibody labeling in materials without Y chromosomes. Improving the quality and transparency of protein-based molecular resources will require the collaborative efforts of producers and consumers. Such efforts will greatly enhance the ability to identify molecular mechanisms of sexual dimorphism in human health and disease. ■

REFERENCES AND NOTES

1. S. K. Lee, *BMB Rep.* **51**, 167 (2018).
2. R. M. Shansky, A. Z. Murphy, *Nat. Neurosci.* **24**, 457 (2021).
3. J. A. Clayton, *FASEB J.* **30**, 519 (2016).
4. A. P. Arnold, *Horm. Behav.* **55**, 570 (2009).
5. B. Sood, R. W. Clemente Fuentes, "Jacobs Syndrome" (StatPearls Publishing, 2022); www.ncbi.nlm.nih.gov/books/NBK557699/.
6. J. P. Dumanski *et al.*, *Am. J. Hum. Genet.* **98**, 1208 (2016).
7. L. A. Forsberg *et al.*, *Nat. Genet.* **46**, 624 (2014).
8. S. Sano *et al.*, *Science* **377**, 292 (2022).
9. A. P. Arnold, *Neurosci. Biobehav. Rev.* **119**, 1 (2020).
10. A. K. Godfrey *et al.*, *Genome Res.* **30**, 860 (2020).
11. H. Shen *et al.*, *Mol. Cell* **82**, 2588 (2022).
12. B. D. Gelfand *et al.*, *MedComm. Futur. Med.* 10.1002/mef2.62 (2023).
13. M. Uhlén *et al.*, *Science* **347**, 1260419 (2015).
14. D. W. Bianchi *et al.*, *Clin. Chem.* **67**, 351 (2021).
15. D. P. Nusinow *et al.*, *Cell* **180**, 387 (2020).

ACKNOWLEDGMENTS

The authors thank C. Gelfand, M. Gelfand, and S. Gelfand for their review. B.D.G. has received funding from the National Eye Institute; the National Heart, Lung, and Blood Institute; the National Institute on Aging of the National Institutes of Health; the BrightFocus Foundation; and the Owens Family Foundation. J.A. has received funding from the National Eye Institute and the National Institute on Aging of the National Institutes of Health.

10.1126/science.ade7187

POLICY FORUM

CARBON MARKETS

Confronting deep uncertainty in the forest carbon industry

Resilience-based and service-focused approaches could reduce contentions and injustices

By **Geoff Wells[1,2,3], Unai Pascual[4,5], Chris Stephenson[6], Casey M. Ryan[1]**

Global momentum on carbon markets has the potential to direct substantial capital toward protecting the world's forests. Yet the billion-dollar forest-carbon-offsetting industry is attracting criticism, in part from doubts about the methods used to measure and causally attribute changes in tree cover and biomass (*1*). Many actors in the industry are thus pursuing increasingly detailed measurement and monitoring of carbon outcomes and risks, under the assumption that this will improve accuracy and offset integrity (*2*). However, mounting scientific evidence (*3*, *4*) implies that many forest landscapes are subject to "deep uncertainty" (*5*), such that claims of high accuracy in assessing carbon change are likely to remain inherently contestable, regardless of the technology or methodology deployed. Further, demands for such accuracy are likely to perpetuate inefficiencies and injustices among carbon suppliers (*6*–*9*). Approaches from other sectors may offer alternative ways forward in the absence of highly accurate measurements of outcomes.

DEEP UNCERTAINTY

Markets thrive on good information. Buyers and regulators in the forest carbon market thus demand this from carbon suppliers, certification standards, monitoring services, and rating agencies. In turn, many of these "supply-side" actors are incentivized to differentiate their offerings through competing claims of superiority in certainty, transparency, and even "truth" (*10*). Supply-side actors survive on their ability to convincingly quantify and causally attribute carbon change.

[1]Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden. [2]Department of Geography, McGill University, Montreal, QC, Canada. [3]School of GeoSciences, University of Edinburgh, Edinburgh, UK. [4]Basque Center for Climate Change, Leioa, Spain. [5]Basque Science Foundation, Ikerbasque, Bilbao, Spain. [6]Independent consultant. Email: geoff.wells@su.se

But the science is far from settled on how (or even whether) it is possible to do this in a robust way. Commodification of forest carbon relies on the principles of additionality, leakage, and permanence. Additionality requires that changes in carbon sequestration or stocks would not have occurred in the counterfactual (e.g., business as usual) scenario (also known as the baseline); controlling for leakage requires that interventions do not increase emissions elsewhere. Permanence requires that carbon stocks remain sequestered on decadal or centennial timescales.

These principles were developed as a political imperative in the 1990s and 2000s to operationalize global agreements on carbon accounting. The assumption that we can meaningfully measure and attribute forest carbon change has subsequently become deeply engrained in market rules and in buyers' expectations (*6*). To assess these principles, the market relies on estimated changes in tree cover and biomass as well as assessments of their causality and risk. The skills and costs of such analytics and monitoring remain the main barrier to entry for projects (*2*, *11*), and investments in new commercial analytical services reach into the tens of millions of dollars (*10*). The perceived accuracy of these analyses by a given project is also often equated with overall carbon offset quality and underpins perennial disagreements about the effectiveness of many projects (*1*).

Yet despite deeply entrenched market expectations about the measurability of forest-carbon change, recent science tells us that these analyses—and thus the commodity of forest-carbon offsets—are likely to remain uncertain. Forest landscapes are complex and dynamic systems in which tree and soil carbon stocks fluctuate in response to diverse interactions between natural (e.g., tree mortality, competition, pests, herbivory, and natural fires) and anthropogenic drivers (e.g., land clearing, wood harvesting, and tree planting). Critically, the relative influences of these drivers vary greatly even across small areas and are usually only partially understood, and only partially observable (*3*).

In ecology, this evidence has seen challenges to theories of climax communities and single stable states (e.g., where undisturbed vegetation is assumed to continually increase in biomass before reaching a predictable and stable maximum) by more dynamic understandings of nonlinearity, hysteresis, and alternative stable states (*4*). The social context of carbon projects adds further uncertainty. Mounting evidence on the complexity of socioeconomic dimensions of land use suggests that past observations of land-use change are not necessarily a good guide to the future and that complex social feedbacks often lead to unintended consequences (*3*). Overall, strong ecological and socioeconomic evidence now exists that forest carbon stocks fluctuate in ways that are hard (and sometimes impossible) to precisely understand, predict, and control.

These advances sit uncomfortably with claims of high certitude about projects or policies that cause forest carbon change. Certainty in additionality and leakage would imply that it is possible to reliably measure biomass change and that one can robustly attribute causality (e.g., through objective counterfactual scenarios and baselines). Permanence would imply that forest biomass can be made to reach something akin to a climax community with stable levels of carbon. But a conservative interpretation of the present science indicates that, no matter the technologies deployed, causal attributions of forest change and expectations of quasi-stable biomass will remain contestable (even where claims are retrospective, e.g., in ex post crediting models). That is, many forest carbon interventions operate in something akin to "deep uncertainty"—a world of unknown unknowns (*5*).

Deep-type uncertainty subsequently points to a more fundamental tension between dominant forest carbon–offset methodologies and advances in the understanding of complexity in socioecological systems. Even if it were possible to robustly assess changes in a singular system-state variable (e.g., tonnes of forest carbon sequestered), complexity science suggests that doing so would (on its own) provide little insight about whether a system has undergone a sustained (e.g., "permanent") regime change (or transformation) to a new (e.g., higher carbon) state (*12*). Such changes are instead more closely associated with the broader stability of the socioecological system (e.g., the enduring environmental, social, and economic factors that encourage or discourage changes in tree biomass) and the ability of an intervention to maneuver



(i.e., its adaptive capacity) to retain a desired state in the face of (unanticipated) shocks and stresses (i.e., its general resilience).

Many carbon schemes do assess "risks" to additionality, leakage, and permanence from specific environmental, social, and economic factors. However, by definition, these risk assessments can only address "known knowns" that can be anticipated and about which one can estimate probabilities and impacts (*5*). They therefore can say little about the likelihood that an intervention will prevail in the face of unanticipated (and sometimes incomputable) shocks and stresses. Thus, not only are quantifications of forest carbon outcomes and associated (specific) risks likely to remain contestable, but subsequent claims of permanence are also likely to be inherently precarious.

COSTS AND INJUSTICES OF CLAIMS TO ACCURACY

In the context of entrenched assumptions and market expectations that forest carbon change is measurable and attributable, supply-side actors have tended to respond to these challenges by investing in ever-more-detailed technical measures for assessing carbon outcomes and risk (e.g., conservative estimates of carbon removals, new modeling and monitoring technologies, offset insurance products, blockchain-enabled transparency measures, and automated risk assessments). Yet evidence of deep uncertainty in forest landscapes suggests that skepticism about carbon measurement and attribution will remain a stubborn burden on the sector.

At present, this burden manifests in different ways. For one, it continues to cause disputes about the worth of different carbon schemes (*1*)—disputes that may simply reflect the preferences of different analysts rather than any fundamental truths about carbon sequestration. Such disputes not only threaten the reputation (and thus the survival) of these individual projects but also diminish market confidence more widely. Persistent uncertainty drives volatility in prices and revenues and has led to an arms race in the sophistication (and often cost) of competing (and still contested) assessment methodologies (*2*). Volatile prices and high compliance costs subsequently remain the main barriers to entry for suppliers, which now renders unviable up to 80% of potential projects in the tropics (*11*).

Questionable assumptions and expectations around certitude also create overburdensome compliance demands, which can in turn fuel unfairness and inefficiency. In the Global South, many carbon suppliers are local organizations, communities, or smallholders from poor areas. Often these suppliers have little capacity to negotiate the rules to which they are subjected and can be pressured to redirect already-scarce resources toward questionably complex and costly compliance measures (*6*, *7*). Suppliers can also be held to external standards of certainty and performance that are virtually unachievable (*8*). Such unfair, costly, and restrictive conditions can subsequently combine to hinder efficiency and innovation (*9*). Ultimately, none of this is good for buyers, suppliers, or the climate.

IDEAS FROM OTHER SECTORS

Despite these challenges, a degree of path dependency and institutional lock-in around market-based approaches to climate change mitigation implies that such approaches are likely to remain economically and politically entrenched—and are likely to remain a key source of capital flows for forest conservation. Within this context, we point to existing approaches from other sectors that could inspire improved approaches to uncertainty in the forest carbon–offset industry.

For one, health, financial, conservation, and other sectors sometimes deploy "value of information" approaches to assess when generating more-detailed data and information is worth the cost, relative to marginal changes in certitude, usability, and user-perceived legitimacy (*13*). Certification organizations, regulators, and other rule-makers in the forest carbon industry could explore similar approaches to more objectively assess what level of complexity is actually warranted in measurement and monitoring forest carbon stocks. More broadly, the industry could explore introducing a fundamental principle of parsimony in methodologies, rules, and assessments—where approaches should be no more complicated than they need to be and should focus on the most uncertain components. Key to this is recognizing that accurately measuring carbon stocks is only a small part of assessing a project's impact and quality.

Another opportunity stems from a potential shift from risk- to resilience-based approaches to assessing the quality of projects, which consider not only known risks but also the capacity to handle the unknowns. Energy, finance, and other sectors approach deep uncertainty by investing in (and assessing) their technical, organizational, social, and economic readiness to persist in the face of unknown shocks and stresses—that is, their adaptive capacity and general resilience (*14*). Although care would need to be taken to not increase (and to preferably reduce) compliance costs, the forest carbon industry could emulate these sectors by moderating its emphasis on the precision of (in any case uncertain) estimated outcomes and instead more closely relate offset quality to general resilience and adaptive capacity of an intervention—for example, a general "stress test" based on the ecological context; the depth and breadth of skills, knowledge, labor, and resources available; and the strength of institutional and governance systems, associated inclusivity, and learning (*12*, *14*). Higher adaptive capacity also usually engenders higher levels of stakeholder consultation and participation, thus also enhancing fairness and legitimacy (*9*).

Somewhat more radically, some operators in these other (noncarbon) sectors address uncertainty by adopting service-focused models that allow trading to proceed in the near-total absence of accurate measurements of outcomes. By using "efforts standards" from contract law, buyers pay suppliers to maximize the delivery of a service, even if there is uncertainty about the precise level of final delivery (*15*). For example, internet service providers agree to undertake specific measures within their control (i.e., "best efforts") to maximize the speed and stability of a service up to a maximum potential level, even if the actual level may be only approximately known and may fluctuate because of events outside of their control (e.g., internet traffic, weather damage to infrastructure).

For forest carbon suppliers, systems theory provides a basis for assuming that although the precise level of a desired system function (e.g., forest carbon sequestration) may not be knowable, it can be assumed to be maximized over time when actors (e.g.,



Forest land near Ashford, Washington, is part of a project that allows the private nonprofit Nisqually Land Trust to sell carbon credits.

forest stewards) use their adaptive capacity (e.g., their land tenure rights and available capital) to maximize the general resilience of a desired system state (e.g., a landscape with more trees) to (often unanticipated) shocks and stresses (e.g., political and economic drivers of deforestation) (*12*). Thus, from a complexity perspective, approaches focused on maximizing adaptive capacity and resilience among carbon suppliers may be more appropriate for managing radical uncertainty rather than pursuing ever-more-detailed (but inherently uncertain) measurements of carbon outcomes.

The exact modalities of how such a service-focused approach could apply in practice would need to be explored, along with how it links to the presently entrenched, commodity-focused (i.e., payments by tonnes of carbon) regulatory and market infrastructure. We do not seek to offer a definitive model here, but where there is a buyer there is a way, and judging by other sectors, a service-focused model may be sufficient for some.

For example, the industry could examine approaches whereby carbon buyers subscribe to a scheme based on an approximate, parsimonious, and externally verified projection of the expected "carbon return" (i.e., removals) over a period. Suppliers could be held to a service agreement that sets out best-effort thresholds (i.e., activities within the limits of their adaptive capacity) for maximizing this carbon sequestration and its resilience. Where these best efforts are met, the supplier would be rewarded and the buyer could make climate claims based on the original expected projection. And this would occur even if the precise level of sequestration remains unknown and may have been higher or lower as a result of exogenous influences. Such approaches may encourage buyers and suppliers to redirect investments toward adaptive capacity and resilience, and supply-side actors may be empowered to be more transparent about what they can actually predict, measure, and control.

Activity-based monitoring is already sometimes used as a precursor to calculating carbon change in the sector, in contrast to approaches that attempt to directly measure carbon change (e.g., outcome-based methods). Complexity science suggests that activity-based approaches on their own may be sufficient (and more appropriate) for maximizing carbon sequestration under deep uncertainty, where further meaningful post hoc assessments of carbon change and attribution are, in any case, often not possible.

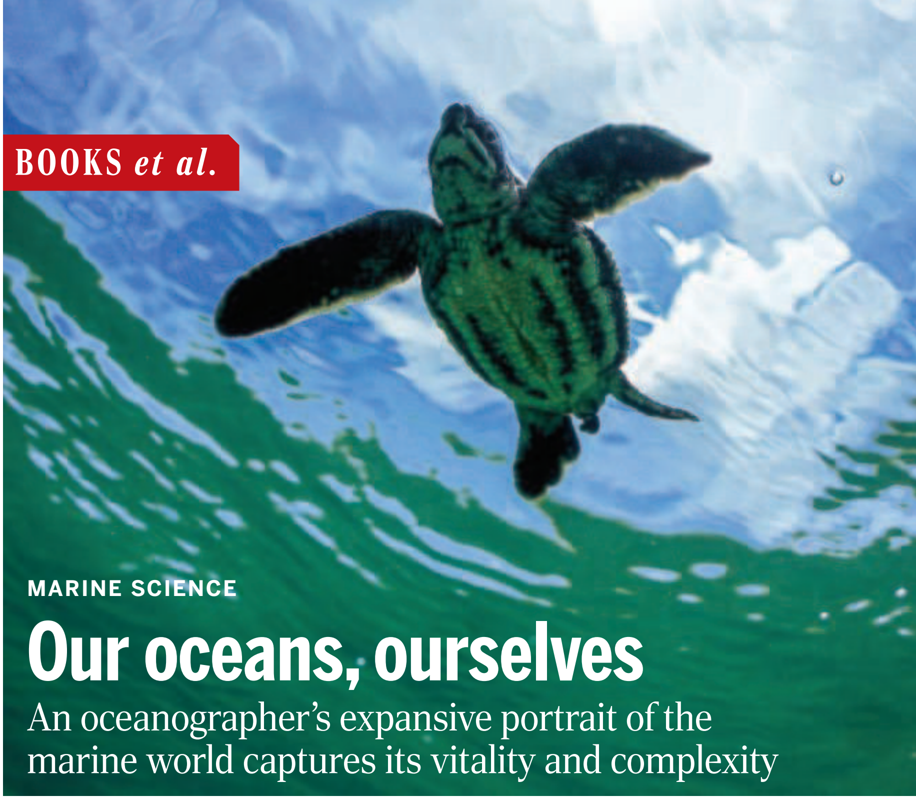
Whatever the models that might emerge, service-focused approaches would likely entail a change in the value proposition to, and the expectations of, buyers and regulators in voluntary and/or compliance markets whereby a buyer's climate claims are based on making something akin to "conditional investments in sustainable and just future landscapes" rather than only purchasing specific amounts of (in any case contestable) carbon offsets. Such a model might also align with the emerging "carbon-contribution" approach to mitigation claims, whereby purchased credits cannot "offset" a buyers' emissions but are instead indicative of their broader commitment to addressing climate change. Under such an approach, approximate quantifications and conditional activity-based payments may suffice.

Generally, we propose that by adopting a complexity-informed perspective and exploring alternative approaches to confronting deep uncertainty, the forest carbon industry (and potentially other ecological markets) may find new pathways for addressing many of the contentions and injustices that presently exist in the sector. ■

REFERENCES AND NOTES

1. Verra, "A non-dogmatic critique of West *et al*. (2023), 'Action needed to make carbon offsets from forest conservation work for climate change mitigation,' *Science* 381: 873–877," *Verra*, 25 August 2023.
2. C. Pan *et al*., *J. For. Res*. **33**, 1109 (2022).
3. P. Meyfroidt *et al*., *Proc. Natl. Acad. Sci. U.S.A*. **119**, e2109217118 (2022).
4. A. Staal *et al*., *Nat. Commun*. **11**, 4978 (2020).
5. E. Wassénius, B. I. Crona, *One Earth* **5**, 35 (2022).
6. A. Gupta, E. Lövbrand, E. Turnhout, M. J. Vijge, *Curr. Opin. Environ. Sustain*. **4**, 726 (2012).
7. F. Fleischman *et al*., *Curr. Opin. Environ. Sustain.* **51**, 7 (2021).
8. R. Carmenta, F. Cammelli, W. Dressler, C. Verbicaro, J. G. Zaehringer, *World Dev*. **145**, 105521 (2021).
9. U. Pascual *et al*., *Bioscience* **64**, 1027 (2014).
10. A. Furey, "We've raised $32M in Series A funding to become a source of truth for carbon markets," *Sylvera Company News Blog*, 25 January 2022; https://www.sylvera.com/blog/series-a-announcement.
11. L. P. Koh, Y. Zeng, T. V. Sarira, K. Siman, *Nat. Commun*. **12**, 1271 (2021).
12. R. Biggs *et al*., *Annu. Rev. Environ. Resour*. **37**, 421 (2012).
13. P. D. Glynn *et al*., *Front. Environ. Sci*. **10**, 805245 (2022).
14. D. Guo, M. Shan, E. K. Owusu, *Buildings* **11**, 464 (2021).
15. K. A. Adams, *Bus. Lawyer* **74**, 677 (2019).

10.1126/science.adh8117



Leatherback sea turtles keep the ocean's salinity bay by constantly expelling salty tears.

MARINE SCIENCE

Our oceans, ourselves

An oceanographer's expansive portrait of the marine world captures its vitality and complexity

By **Todd L. Capson**

In her new book, *The Blue Machine*, physicist and oceanographer Helen Czerski describes a dynamic "ocean engine"—a tightly woven system that converts sunlight into movement and life—writing on scales ranging from bubbles to the Gulf Stream and from krill to blue whales and highlighting the ocean's vital connections to human history and culture. Her description of the ocean, including its complexity and diversity, is vast and comprehensive. Yet *The Blue Machine* is a point of departure, a map for further exploration. Not since reading *The Diversity of Life* by E. O. Wilson have I read a book as timely, salient, and informative.

Czerski explains a broad range of important phenomena, properties, and concepts and provides examples that are engaging and illuminating. The following summaries illustrate her approach. She describes, for instance, how the sunlit upper layer of the ocean provides heat energy to the ocean and sculpts global weather patterns. Underneath are thicker layers of varying salinity and temperature, typical of ocean basins. She renders the abstract concrete, describing the battle between Mark Antony and Octavian in 31 BCE in which Antony's ships were possibly stymied by "dead water"—internal waves in denser, saltier water below that impede forward motion. Octavian would go on to establish the Roman Empire.

The Blue Machine illustrates the critical role of density to the physical structure of the ocean with a description of the Denmark Strait Overflow, where 3 million m^3 per second of dense, cold, salty Arctic water flows underneath the Atlantic. This feature is an important component of the thermohaline circulation system that transports warm water from the tropics to the poles. Czerski explains how salinity has major repercussions on the physical structure and machinery of the ocean and its inhabitants, such as the leatherback turtle, which copes with a daily salt intake of 10 kg by crying 8 liters of salty tears every hour. She describes the overall shunting of energy from equatorial regions to the poles and how polar regions play critical roles for the flow of energy from Earth into space. She explains what the North and South Poles have in common and how they differ.



The Blue Machine: How the Ocean Works
Helen Czerski
Norton, 2023. 464 pp.

Czerski writes of the travails and joys of being an ocean scientist, including an expedition to study the exchange of gases between the ocean and atmosphere in the middle of a storm in the North Atlantic. She writes of the importance of collecting data in the field that will keep computer models from "stray[ing] into the world of computer games."

Describing her initial exposure to ocean science as a postdoc, after obtaining three physics degrees, Czerski reflects on her exasperation that "no one had ever mentioned the ocean." Her reaction reminded me of the response of Mireya Moscos, then president of Panama, to a presentation on why the Coiba National Park was worthy of protection ("Why didn't anyone tell me?") (*1*).

Czerski writes about society's tendency to see the ocean as "away," a solution to the Great Stink of London in 1850, when excrement was pumped into the North Sea until returning tides and fouled fisheries forced a better solution. Nature, on the other hand, puts waste to good use, as demonstrated by the blue whales in the Southern Ocean, whose generous clouds of excrement bring nutrients into sunlit surface waters to be recycled by the splendidly named krill *Euphausia superba*.

The Blue Machine provides important context to help readers better understand humanity's impacts on the ocean. We have demolished the balance of the global carbon budget by burning fossil fuels, resulting in what Nicolas Gruber of ETH Zürich calls the "triple whammy" of ocean acidification, warming, and deoxygenation (*2*). We now observe unprecedented ocean surface temperatures that supercharge storms and alter global weather patterns, pervasive hypoxia on global coral reefs, and warnings of a forthcoming collapse of the Atlantic Meridional Overturning Circulation, among other challenges. The traditional Polynesian navigation techniques that Czerski describes—underpinned by teamwork, humility, and a profound understanding of the natural world—provide a model for addressing these threats, a stark contrast to the hubris of interventions such as geoengineering.

My only reservation about the book is Czerski's use of metaphors and other literary embellishments that sometimes get in the way of her otherwise clear and descriptive prose—although she is entitled to some creative license for having written a tome that is both substantive and accessible. Ultimately, *The Blue Machine* is a major contribution to our shared responsibility of educating, inspiring, and informing societal values about the ocean. ■

The reviewer is at the Institut de physique du globe de Paris, Université Paris Cité, 75005 Paris, France, and Parley for the Oceans, New York, NY 10013, USA. Email: capson@ipgp.fr

REFERENCES AND NOTES

1. Legislation was adopted for the protection of Coiba National Park in 2004, and it was inscribed to the list of UNESCO World Heritage Sites in 2005. See T. L. Capson, in *Global Resource Scarcity: Catalyst for Conflict or Cooperation?*, M. C. Dawson, C. Rosin, N. Wald, Eds. (Routledge, 2017), chap. 7.
2. N. Gruber, *Philos. R. Soc. Lond. Ser. A* **369**, 1980 (2011).

10.1126/science.adj7198

SCIENCE LIVES

The mRNA enthusiast's memoirs

A molecular biologist documents how her passion for the messenger molecule paid off

By **Vijaysree Venkatraman**

Breaking Through: My Life in Science
Katalin Karikó
Crown, 2023.
336 pp.



One day in 1997, University of Pennsylvania molecular biologist Katalin (Kati) Karikó met a new hire at the photocopier who would change the trajectory of her research forever. At 42, the scientist had made little headway with her big idea that messenger RNA (mRNA)—the ephemeral molecule that instructs cells to make proteins—could be commandeered to help the body fight disease. Her grant proposals to study the therapeutic potential of mRNA continued to get rejected. She had no budget, no staff, and very little prestige.

Karikó introduced herself to the "photocopier-hogging immunologist," Drew Weissman, and asked about his work. He planned to create vaccines against infectious diseases and discussed ways to deliver an antigen into cells to trigger the body's own immune response. At once, Karikó saw another prospective role for her pet molecule, and a longstanding collaboration was born that laid a strong foundation for the mRNA-based vaccines that proved vital during the COVID-19 pandemic. In *Breaking Through*, Karikó's page-turning memoir, the talented and tenacious researcher recounts the origins and evolution of her scientific journey.

Karikó grew up in a small town in Hungary, the daughter of a skilled butcher. There, a high school biology teacher introduced her to endocrinologist Hans Selye's *The Stress of Life*—a book that would resonate with her both as an introduction to scientific thinking and as a guide to living well. Nature "rarely replies to questions unless they are put to her in the form of experiments, to which she can say yes or no," wrote Selye in that book in 1956.

Selye's observations helped Karikó realize that she wanted to spend her life designing experiments that would contribute to human knowledge. Her parents encouraged her to become a *kutató*—Hungarian for "searcher"—or, in English, a researcher.

She earned her PhD from the prestigious Biological Research Centre in Szeged, Hungary, where she began her exploratory research on medicinal applications for RNA. When her funding was pulled, she found a postdoc position in Philadelphia. She moved to the United States with her husband and toddler, Susan, with the family's meager savings sewn into the little girl's teddy bear.



Karikó's hunch that mRNA could be used to fight disease proved correct.

"Do not blame," wrote Selye. "Focus on what you can control. Transform bad stress into good stress." This advice would serve Karikó well over the course of her nearly four-decade-long academic career, a period, she writes, in which she was belittled, ignored, demoted, and even threatened with deportation when her research did not win government grants or private funding.

Still, nothing upset her as much as one early lab finding: Synthetic mRNA triggers an inflammatory immune response in cells. An mRNA vaccine might still have been possible, Karikó and Weissman reasoned, but mRNA would be useless as therapeutic medicine until they figured out a way to keep it from activating the immune system.

In 2005, when the pair reported that a modified mRNA produced a noninflammatory immune response in cells, their landmark finding (*1*), which would eventually be licensed by biotech companies, went unheralded. In 2013, in a rude reminder that lab space is often allocated on the basis of the funding a researcher brings in, Karikó was unceremoniously displaced. "I showed up to my lab in the neurosurgery department and found my belongings in the hallway," she writes. She was moved to a tiny room that was not conducive for working with mRNA. This was the final straw.

That year, Karikó received job offers from two companies that had licensed the modified mRNA technology: BioNTech in Germany and Moderna in the United States. When she packed her bags and left temporarily for Germany, her daughter—by now a two-time Olympic gold medalist—and husband stayed behind.

Karikó compares her science with her daughter's rowing. The crew rows backward, she observes, without a view of the finish line. Just as she has done in the lab, they must trust that if they give it their all, their hard work will bring them to the right destination.

Karikó's discovery found a meaningful application during the COVID-19 pandemic, but the story of synthetic mRNA technology will not end here. Scientists are studying its therapeutic potential for various cancers, cystic fibrosis, and rare metabolic disorders, as well as for vaccines against other infectious diseases, Karikó writes. In the next decade, she predicts, there will be an explosion of new mRNA therapies and vaccines. Her moving memoir offers a fascinating glimpse into the dead ends and hard work that led up to this important discovery, while highlighting critical shortcomings in the scientific reward system. ■

The reviewer is a science journalist in Boston, MA 02144, USA. Email: v.vijaysree@gmail.com

PHOTO: BLOOMBERG VIA GETTY IMAGES

REFERENCES AND NOTES

1. K. Karikó, M. Buckstein, H. Ni, D. Weissman, *Immunity* **23**, 165 (2005).

10.1126/science.adj6123

INSIGHTS

LETTERS



NEXTGEN VOICES

Historic introductions

We asked young scientists this question: **If you could introduce any two scientists, regardless of where and when they lived, whom would you choose, and how would their collaboration change the course of history?** Read a selection of the responses here. Follow NextGen Voices on social media with hashtag #NextGenSci.

Improved medicine

Hungarian-American biochemist Katalin Karikó has contributed to the mRNA technology that made the COVID-19 vaccine possible. French virologist Françoise Barré-Sinoussi first identified HIV (along with Luc Montagnier). Both scientists are alive today, but I would have liked to introduce them in 1983, when Barré-Sinoussi first made her HIV discovery. The collaboration between these two great minds decades ago could have changed the course of multiple pandemics.

Keren Bismuth
Paris, France. Email: bismuthk2@gmail.com

French physicist Marie Curie (born in Poland) lost her life in 1934 to aplastic anemia caused by radiation exposure. German immunologist Paul Ehrlich, who lived at the same time, first identified the cause of aplastic anemia. I would have liked to introduce these two scientists in time for Ehrlich's knowledge to prevent Curie's death. Together, they could have educated the world about the negative effects of radiation, and perhaps even worked on a cure for aplastic anemia.

Vandana Sharma
Department of Hematology, All India Institute of Medical Sciences, New Delhi, New Delhi, India. Email: sharmavandana.phd@gmail.com

Santiago Ramón y Cajal, born in Spain in 1852, is said to have been so brilliant that simply by looking through early 1900s microscopes, he was able to intuit the role of specific neurons. What if Cajal had been able to meet one of today's pioneers in the intersection of technology and neuroscience? French neuroscientist Grégoire Courtine's recent advancements include developing an integrated brain–computer interface to recover walking after spinal cord injury.



Cajal's discoveries were limited only by the tools available to him. If paired with a researcher capable of driving technology forward, it's hard to envision what problems in neuroscience couldn't be solved.

Jackson Ross Powell
Vagelos Molecular Life Sciences Program, University of Pennsylvania, Philadelphia, PA 19104, USA. Email: jrp24@sas.upenn.edu

A collaboration between Hua Tuo, a second-century Chinese surgeon, and Leonardo da Vinci, a 15th-century Italian architect and anatomist, would combine ancient Chinese medical wisdom with Renaissance artistic brilliance. Hua Tuo could impart the essence of traditional Chinese medicine, including acupuncture and herbal remedies, to da Vinci, allowing him to portray the human body in a more holistic way. In return, Leonardo's meticulous anatomical studies and innovative scientific approaches could revolutionize Hua Tuo's medical practice, enhancing surgical techniques and diagnostics. Their collaboration might have ignited a comprehensive healthcare approach that blended Eastern and Western wisdom and altered how the human body is understood, treated, and depicted.

Haoxian Tang
Shantou University Medical College, Shantou, Guangdong, China. Email: 19hxtang@stu.edu.cn

Leonardo da Vinci, 15th-century Italian painter, engineer, architect, and inventor, and Richard Feynman, 20th-century US theoretical physicist, could have worked together to revolutionize medical technology. Da Vinci's inventions include "Leonardo's robot," a functional automaton, which may have served as the inspiration for today's widely used robotic surgical systems. If da Vinci had met Richard Feynman, the father of nanotechnology, perhaps they could have created "Leonardo's nano-robot," a micromachine used to remove bacteria and viruses, repair blood vessels, and kill tumor cells accurately in vivo.

Bo Cao
Core Research Laboratory, The Second Affiliated Hospital, School of Medicine, Xi'an Jiaotong University, Xi'an, Shaanxi, China.
Email: bocao@vip.qq.com

German microbiologist Robert Koch, born in 1843, discovered the microbes that cause anthrax and tuberculosis and pioneered bacteriology. Scottish doctor Alexander Fleming, born in 1881, discovered penicillin, the first all-purpose antibiotic. If these two researchers had worked together, they could have made great strides in topics such as the traits of pathogens, the mechanics of transmission, and antibiotic responses, leading to a better understanding of infectious diseases and more effective antibiotics.

Jingtao Huang
Shantou University Medical College, Guangdong, China. Email: 19jthuang@stu.edu.cn

Robert Hooke, a 17th-century English physicist, developed the first microscope. This tool would have enabled ancient Greek anatomist Herophilus, who lived in the 3rd century BCE, to extend his studies beyond macroscopic explorations of the human body. Having described the retina and optic nerve, Hooke's microscope would have enabled Herophilus to discover, for instance, the many cell types composing these structures. Herophilus's embrace of the scientific method and experimentation would have led to huge strides in demystifying the human body and its functionality. Instead of being chastised at the time as a useless endeavor, the study of anatomy would have flourished, fundamentally transforming the course of medicine and human health.

Rishi Jai Patel
Department of Chemistry, University of Pennsylvania, Philadelphia, PA, USA.
Email: ripatel@sas.upenn.edu

Advanced technology

Leonardo da Vinci, the 15th-century Italian polymath known for his engineering, architectural, and artistic prowess, and Nikola Tesla, a Serbian electrical and mechanical engineer born in 1856 in the part of the Austrian Empire that is now Croatia, possessed complementary knowledge, skills, and insights. Tesla's understanding of electricity might have been the key to making da Vinci's sophisticated designs feasible. Conversely, da Vinci's wide variety of creative designs could have inspired Tesla to apply his electricity expertise to other fields. If the two of them had collaborated, helicopters and submarines might have been invented earlier, altering the course of history.

Paulo Bezerra
Department of Engineering and Technology, Universidade Federal Rural do Semi-Árido, Pau dos Ferros, RN, Brazil.
Email: paulo.bezerra@ufersa.edu.br

Ada Lovelace, born in England in 1815, was renowned for her contributions to early computer science. Her legacy as the world's first computer programmer is primarily based on her groundbreaking work with Charles Babbage's analytical engine—an innovative early mechanical general-purpose computer. Alan Turing, born in England in the 20th century, specialized in computer science, mathematics, and cryptography. He is celebrated for contributions to early computing, artificial intelligence, and cryptography during World War II. Their collaboration might have accelerated the development of early computers and algorithms, potentially giving rise to artificial intelligence much earlier in history. Programmable computers in the 19th century could have reshaped communication, cryptography, and automation as well as influenced historical events such as wars.

Xuan Zhang
Shantou University Medical College, Shantou, Guangdong, China. Email: 19xzhang@stu.edu.cn

Geoffrey Hinton, a modern-day British-Canadian cognitive scientist and computer scientist, is a pioneer in artificial intelligence (AI). The use of AI is growing, but because of the complex mathematical relationships on which it depends, its internal workings remain largely a black box. Srinivasa Ramanujan, an Indian mathematician who lived at the turn of the 20th century, possessed an uncanny ability to invent mathematical tools to model difficult real-world problems. If they collaborated, Ramanujan's skill could illuminate the tangled associations inside AI neural networks. Meanwhile, Hinton's extensive experience in building AI algorithms and systems would be vital to implement Ramanujan's mathematical theories as enhanced AI. Together, Ramanujan's unique brilliance in crafting abstract mathematics and Hinton's technical creativity could unravel the cryptic core of AI. Their collaboration might produce AI that is more interpretable, controllable, trustworthy, and aligned with human values and ethics.

Qianjun Wen
The Affiliated Hospital of Guizhou Medical University, Guiyang City, Guizhou Province, China. Email: wqjtmmu@126.com

Greater equity

I would introduce 19th century English naturalist Charles Darwin and modern-day Brazilian psychologist Deisy das Graças de Souza, who has explored how behavior patterns affect reading and writing skills. Both Darwin's theory of evolution and the study of behavior evolution emphasize the role of the environment—Darwin focused on species' adaptations over generations and de Souza studies how an individual's behavior changes as a result of environmental consequences. Marrying Darwin's evolutionary principles with de Souza's expertise in learning might have revolutionized our understanding of how to arrange the environment to promote learning and positive change at both individual and cultural levels.

Fernanda S. Oda
Department of Applied Behavioral Science, University of Kansas, Lawrence, KS, USA. Email: oda@ku.edu

Introducing 20th-century US virologist Jonas Salk, the ingenious mind behind the polio vaccine, and recently deceased physician Paul Farmer, the esteemed medical anthropologist dedicated to addressing healthcare disparities, would produce a promising collaboration. Salk's selfless decision not to patent his vaccine perfectly aligns with Farmer's unwavering commitment to equitable healthcare access. By combining Salk's breakthroughs with Farmer's sociomedical perspective, they would revolutionize our approach to tackling epidemics and ensure global health for all. Salk would learn about the complex social determinants affecting healthcare access from Farmer, and Farmer could gain insights into cutting-edge scientific advancements from Salk. Their united vision would inspire a paradigm shift, emphasizing humanitarianism over profit, ensuring that lifesaving treatments and preventive measures are accessible to everyone, regardless of their background.

Olesia Verstiuk
University of Nicosia Medical School, Nicosia, Cyprus. Email: verstiuk.o@live.unic.ac.cy

Abu Ali al Hasan ibn al-Haytham (known in the West as Alhazen), was a 10th-century mathematician and optician who lived in what is now Iraq. Introducing him to 17th-century English physicist Sir Isaac Newton would facilitate an unparalleled exchange of knowledge and ideas from two very different parts of the world, leading to revolutionary discoveries in the field of light and optics. More importantly, this meeting would promote cultural exchange between some of the brightest minds of the Islamic world and the West, raising the profile of the contributions of Muslim scientists to the world's intellectual heritage. The rich legacy of the Islamic Golden Age (often referred to as the medieval or Dark Ages of the West) is frequently overshadowed in Western historical narratives, so this encounter could pave the way for greater cross-cultural scientific exchange and mutual intellectual appreciation.

Muhammad Shehryar Khan
Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. Email: mskhan@mit.edu

What would have happened if the first public warning that climate change was an existential challenge for humanity had come before Edward Teller's remarks in 1959? A collaboration between Swedish physical chemist Svante Arrhenius, born in 1859, and modern-day US meteorologist Syukuro Manabe could have resulted in an earlier physical model of Earth's climate system that allowed them to quantify the impact of human-induced climate change. A warning of the climate crisis at the start of the 20th century might have given humanity invaluable time to transition from fossil fuel-based energy systems to net-zero emission ones. This scientific collaboration might have reduced the likelihood of experiencing catastrophic consequences of climate change.

Edgar Virgüez
Department of Global Ecology, Carnegie Institution for Science, Stanford, CA, USA. Email: evirguez@carnegiescience.edu

David Harold Blackwell, US statistician and mathematician born in 1919, and George Washington Carver, US agricultural scientist born around 1864, could have collaborated to increase food security. Blackwell's mathematical theory would help Carver study the genetic materials and characteristics of crops more accurately, so as to breed crops that meet people's needs and improve farming methods. Carver's agricultural production data and experience could provide valuable practical materials for Blackwell's research on statistics and decision-making.

Yuan Zhi
School of Economics, Guizhou University, Guiyang, Guizhou, China. Email: yzhi@gzu.edu.cn

When 18th-century English immunologist Edward Jenner arrives—1600 years before his time—at the Asclepion (healing temple) of the ancient Greek city of Pergamon, Galen takes no notice. Preoccupied with his theory of the four humors and dissections of Barbary macaques, Galen is uninterested in milkmaids' lesions. But in 166 AD, Jenner uses cowpox to inoculate a young boy to the Antonine Plague (known today as smallpox). Galen, great student and synthesizer, engages Jenner, soon recording Jenner's method in one of his medical treatises. Galen's article is transcribed and traded, and smallpox is eradicated by 300 AD (17 centuries before 1980). Other vaccinations follow. When Christopher Columbus arrives in the "New World," his men and the Indigenous peoples exchange minor infections, but there is no epidemic. Over the next few centuries, as settlers push westward and wars break out, many Native American tribes, at full strength given the lack of new diseases, are able to repel the invaders. The United States, as it looks today, is never established. This journal is never founded, and this question is never asked. But hundreds of millions of lives are spared.

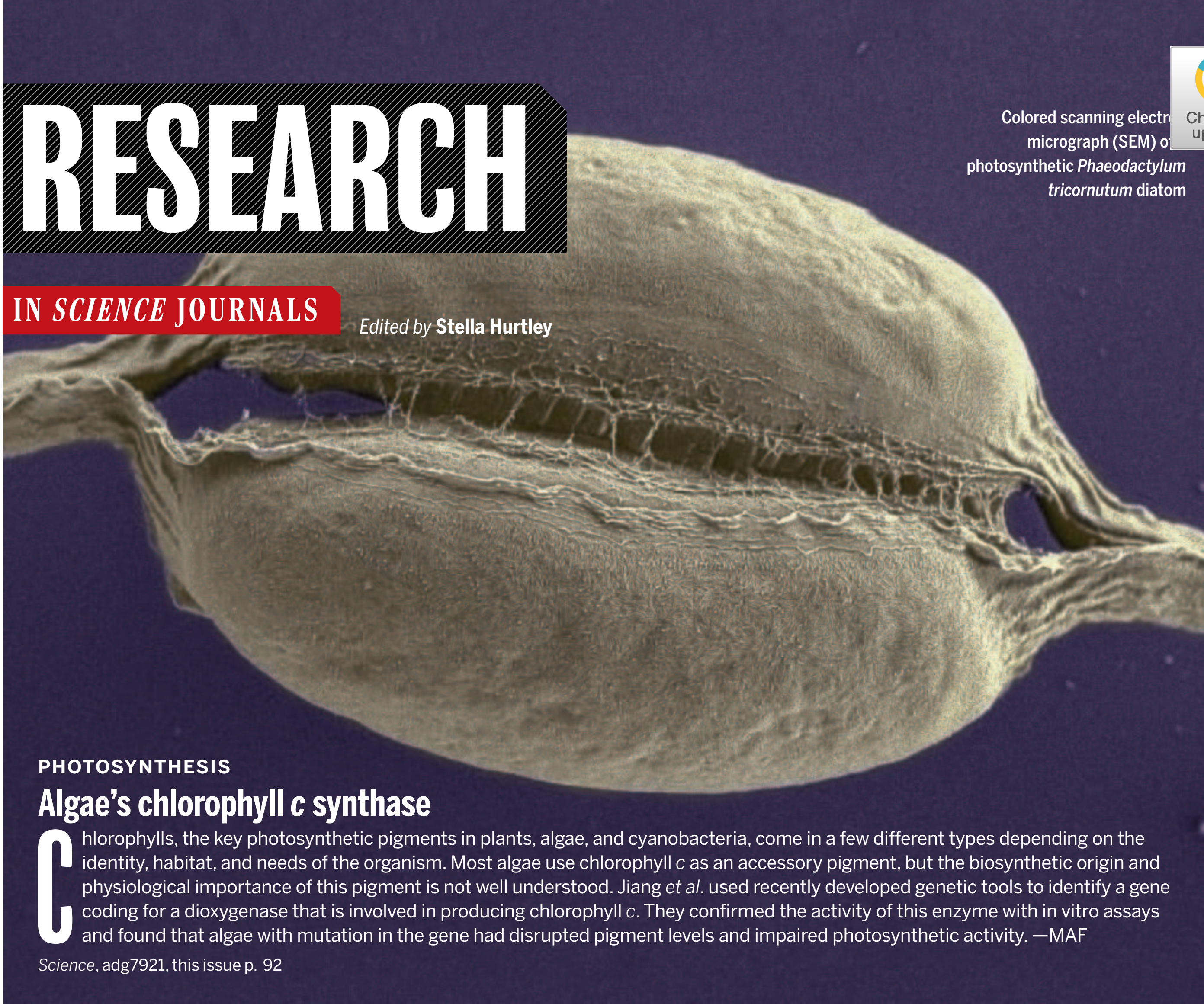
John M. Dedyo
Yale University, New Haven, CT, USA. Email: johnny.dedyo@yale.edu

10.1126/science.adk8769

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**



Colored scanning electron micrograph (SEM) of photosynthetic *Phaeodactylum tricornutum* diatom

PHOTOSYNTHESIS

Algae's chlorophyll *c* synthase

Chlorophylls, the key photosynthetic pigments in plants, algae, and cyanobacteria, come in a few different types depending on the identity, habitat, and needs of the organism. Most algae use chlorophyll *c* as an accessory pigment, but the biosynthetic origin and physiological importance of this pigment is not well understood. Jiang *et al*. used recently developed genetic tools to identify a gene coding for a dioxygenase that is involved in producing chlorophyll *c*. They confirmed the activity of this enzyme with in vitro assays and found that algae with mutation in the gene had disrupted pigment levels and impaired photosynthetic activity. —MAF

Science, adg7921, this issue p. 92

GPCR SIGNALING

How neutrophils decide on a destination

Neutrophils are simultaneously exposed to multiple chemoattractants that bind to $G\alpha_i$ protein–activating receptors. Lundgren *et al.* investigated how neutrophils rank chemoattractants that activate a common signaling pathway in human neutrophils and a neutrophil cell line. Signaling induced by chemoattractants typically released from sites of infection or damage lasted longer than that induced by lower-priority chemoattractants. The rapid attenuation of signaling downstream of a low-priority chemoattractant receptor required the phosphorylation of the C-terminal tail of the receptor. —WW

Sci. Signal. (2023) 10.1126/scisignal.add1845

BIOCHEMISTRY

Radical at the heart

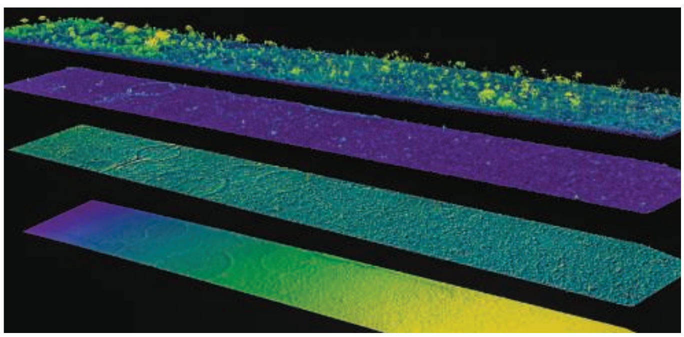
To synthesize deoxyribonucleotides, organisms rely on a radical reaction catalyzed by a family of enzymes called ribonucleotide reductases. Generating the required radical and transferring it to the correct residues within the enzyme active site is a major biochemical undertaking. Lebrette *et al*. have visualized the structure of one of the radical generation and storage subunits of this enzyme with an intact radical present in its core. The water molecules and hydrogen bonds around the dihydroxyphenylalanine radical are restructured relative to states where the radical is lost or transported out of the protein. Observing this structure is only possible using serial femtosecond crystallography methods that avoid x-ray damage entirely. —MAF

Science, adh8160, this issue p. 109

ANTHROPOLOGY

Hidden societies

Indigenous societies have lived in the Amazon for at least 12,000 years. Finding evidence of these societies, however, has been greatly hampered by the density of the forest in Amazonia. Peripato *et al*. used LIDAR (light detection and ranging) surveys to identify more than 20 previously unidentified developments and then modeled the occurrence of others across the Amazon. The authors predict that between 10,000 and 24,000 ancient earthworks are waiting to be discovered. Sampling of some of the LIDAR transects revealed a



LIDAR data reveal earthworks (rings) beneath the forest canopy in Amazonia.

consistent set of domesticated tree species associated with the developments, suggesting active forestry practices among these societies. —SNV

Science, ade2541, this issue p. 103

QUANTUM INFORMATION

Building a qubit system atom by atom

Qubits implemented using solid-state spins are easy to manipulate, but such architectures remain tricky to scale up. Wang *et al.* used a scanning tunneling microscope (STM) to engineer an atomic-scale multi-qubit system consisting of a sensor qubit and up to two remote qubits. The sensor qubit, a titanium atom on a bilayer magnesium oxide film, was directly controlled by the STM tip, whereas the remote qubits were each placed next to an iron atom and read out using the sensor qubit. The researchers demonstrated single- and multi-qubit gates. The size and performance of the setup may be further improved by optimizing the choice of materials and spin species. —JS

Science, ade5050, this issue p. 87

CORONAVIRUS

A series of unfortunate events

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic was marked by waves of new strains of virus differing in virulence and immune reactivity. The advent of each new variant of concern brought more human casualties and waves of onerous quarantine measures. To map the evolutionary trajectory of the variants, Wilks *et al*. obtained sera from people who had been vaccinated or infected with a range of variants of concern and applied antigenic cartography to visualize structural changes in the virus. The authors observed changes in immunodominance and immune escape depending on the variant that had infected the patient or after vaccination. Such analysis has implications for variant risk assessment and for selecting the next candidate vaccine strains that will confer the highest protection. —CA

Science, adj0070, this issue p. 68

VACCINES

An mRNA vaccine for mpox

The recent mpox virus (MPXV) outbreak has highlighted the need for high-quality and safe Orthopoxvirus vaccines. The vaccine administered in response to the outbreak, a modified vaccinia ankara (MVA), was effective at reducing mpox severity and transmission; however, a next-generation vaccine may have performed even better. Freyn *et al*. tested such a candidate vaccine, a messenger RNA (mRNA) vaccine encoding four highly conserved MPXV antigens. The mRNA vaccine performed as well as or better than an MVA comparator in terms of eliciting immune responses and protecting against lethal infection in mice. —CSM

Sci. Transl. Med. (2023) 10.1126/scitranslmed.adg3540

IMMUNOLOGY

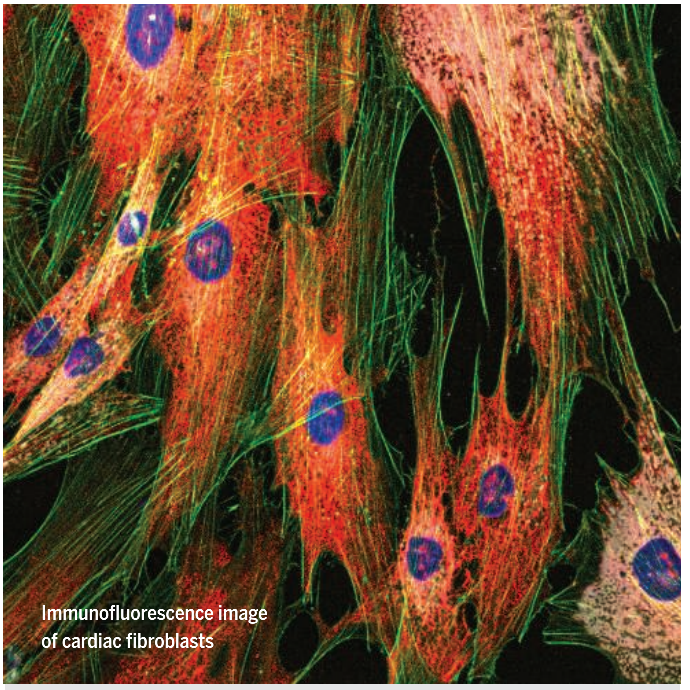
Wash our hands of it?

The hygiene hypothesis proposes that early-life exposure to microbes protects against the development of allergies. However, a causal link between early-life infections and allergies is lacking. To uncover the impact of lifelong microbial exposure on the development of allergic inflammation, Ma *et al*. compared experimental wildling mice, which have a representative breadth of naturally occurring pathogens, with specific pathogen–free mice. Surprisingly, when challenged with allergens, wildlings developed robust signs of pathological inflammation and allergic responses. Thus, increased microbial biodiversity in wildlings does not protect against allergic inflammation, in contrast to predictions of the hygiene hypothesis. —HMI

Sci. Immunol. (2023). 10.1126/sciimmunol.adf7702

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**



Immunofluorescence image of cardiac fibroblasts

CARDIOLOGY

A CHIP off the old heart

Clonal hematopoiesis of indeterminate potential, or CHIP, is the buildup of potentially oncogenic mutations in hematopoietic stem cells, and its risk greatly increases with age. It is not a malignant condition itself, but it does predispose patients to hematological cancers and to multiple types of cardiovascular disease, increasing the risk of death. Abplanalp *et al*. and Rauch *et al*. examined the effects of common CHIP mutations on monocytes and macrophages, blood cells that play key roles in the process of inflammation. The researchers linked the presence of CHIP mutations in these cells to pro-inflammatory activity, helping to explain how CHIP contributes to both heart failure and atherosclerosis. —YN

Nat. Cardiovasc. Res. (2023) 10.1038/s44161-023-00322-x, 10.1038/s44161-023-00326-7

IMMUNOLOGY

New roles for an immune stalwart

Group 2 innate lymphoid cells (ILC2s) tend to reside in tissues where pathogens enter the body, and they contribute to immunity against parasite infection. Like other types of immune cells, excessive proinflammatory activity of ILC2s can contribute to disease. Cui *et al*. found that CD45, a transmembrane tyrosine phosphatase that is important for the development and activation of T cells, plays a role in restraining ILC2s. Deletion of CD45 in ILC2s increased cell numbers and maturation under steady-state conditions and triggered their hyperactivation during immune responses. This was associated with exacerbated lung inflammation and pathology in mouse models of asthma and pulmonary fibrosis. The results not only show the regulatory

PHOTO: MARIANA SHUMLIAKIVSKA

ALSO IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

MICROBIOLOGY

Protection from friendly fire

Mammalian cells use guard mechanisms to monitor their functional pathways for interference by pathogens. Infection causes the production of the inflammatory cytokine interferon-γ (IFN-γ), which triggers the expression of hundreds of IFN-stimulated-genes, including the kinase PIM1 and GBP1, a membrane-perturbing GTPase. Fisch *et al*. identified a guard mechanism whereby PIM1 phosphorylates GBP1 and subjects it to sequestration by a 14-3-3 protein. In human macrophages, this mechanism was found to prevent GBP1 activity from causing Golgi fragmentation and cell death. Pathogens can interfere with IFN-γ signaling and thereby potentially escape immune detection. However, when this signaling is inhibited, short-lived PIM1 is degraded, which allows GBP1 to control pathogen growth. These findings suggest a model of IFN-γ–dependent protection of uninfected bystander cells against self-inflicted innate immune damage. —SMH

Science, adg2253, this issue p. 67

MATERIALS SCIENCE

Moving dislocations in diamond

The speed of sound is often a limit of a sort on how fast something can move through a system. For dislocation motions that play a large part in plastic deformation, the limits are very poorly constrained from experiments. Katagiri *et al*. used x-ray radiography of shocked single-crystal diamond to track dislocation motion during plastic deformation (see the Perspective by Knudson and Seagle). They found that the dislocation speed can be higher than the bulk sound speed of the material. This observation is important for refining deformation models at extreme conditions because these ultrafast dislocation motions have previously only been predicted with theory. —BG

Science, adh5563, this issue p. 69;
see also adk4420, p. 37

HUMAN EVOLUTION

Expanding the evidence

Traditionally, researchers believed that humans arrived in North America around 16,000 to 13,000 years ago. Recently, however, evidence has accumulated supporting a much earlier date. In 2021, fossilized footprints from White Sands National Park in New Mexico were dated to between 20,000 and 23,000 years ago, providing key evidence for earlier occupation, although this finding was controversial. Pigati *et al*. returned to the White Sands footprints and obtained new dates from multiple, highly reliable sources (see the Perspective by Philippsen). They, too, resolved dates of 20,000 to 23,000 years ago, reconfirming that humans were present far south of the ice sheets during the Last Glacial Maximum. —SNV

Science, adh5007, this issue p. 73;
see also adk3075, p. 36

NEUROSCIENCE

Preparing for motherhood

Motherhood leads to pronounced behavioral changes in many species, such as altered feeding routines and increased aggressivity. How does pregnancy prepare females for such future behavioral needs? Ammari *et al*. discovered that the hormonal milieu of pregnancy remodels a distinct population of hypothalamic neurons in mice (see the Perspective by McCarthy) that mediates the onset of parental behavior before giving birth. Sensing of estradiol and progesterone by galanin-expressing neurons in the medial preoptic area is necessary for this behavioral change. Therefore, hormone-mediated neuronal modifications lead to an increased selectivity for pup stimuli, thus anticipating future parenting behaviors. —PRS

Science, adi0576, this issue p. 76;
see also adk2495, p. 33

2D MATERIALS

Landau levels in a twisted system

In a perpendicular magnetic field, electrons in two-dimensional (2D) materials occupy discrete energy levels known as Landau levels. These levels are described by the so-called Onsager relation. Slot *et al*. used scanning tunneling spectroscopy to measure the Landau levels in twisted double-bilayer graphene with an intermediate twist angle of 1.74° between the two bilayers. By comparing their results against the theoretical expectations, the researchers were able to measure deviations from the standard form of the Onsager relation stemming from the orbital magnetism in the system. —JS

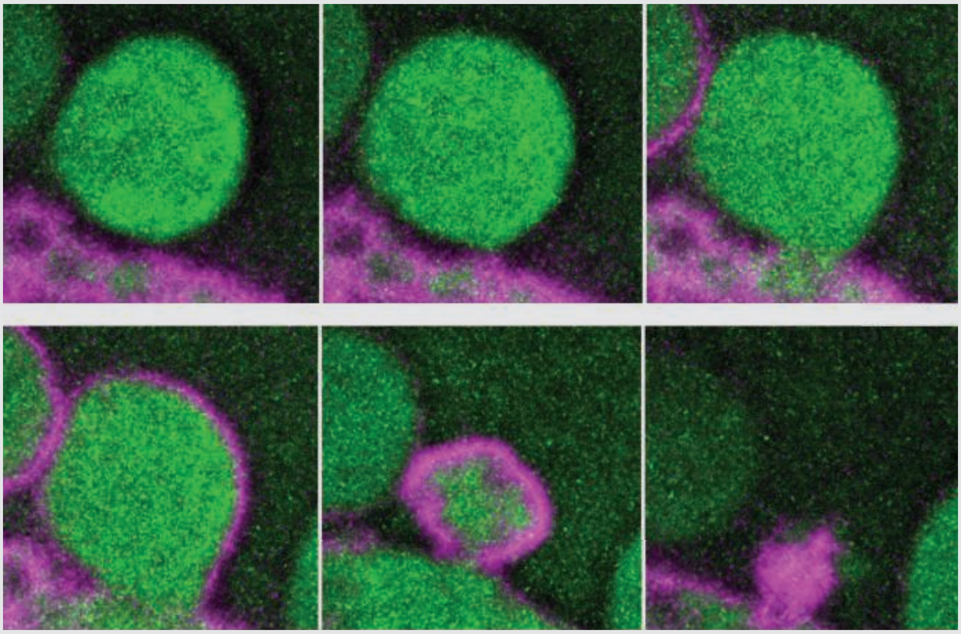
Science, adf2040, this issue p. 81

CATALYSIS

Oscillations in hydrocarbon synthesis

Reaction temperature and product formation oscillate during the hydrogenation of carbon monoxide to form hydrocarbons (Fischer-Tropsch synthesis) over cobalt–ceria catalysts. Zhang *et al*. observed long-duration oscillations of several degrees Celsius with a period of several minutes for a 1:1 reactant feed over cobalt–cerium oxide powder catalyst operating at 220°C, as well as changes in product formation and reaction rates (see the Perspective by Niemantsverdriet and Westrate). A kinetic model including a carbon monoxide insertion step, thermal activation of carbon–oxygen bonds, and periodic temperature forcing could account for the observations. —PDS

Science, adh8463, this issue p. 99;
see also adk5831, p. 35



CELL BIOLOGY

Exocytosis by membrane crumpling

Exocrine secretion is less well understood than neurotransmitter release. Exocrine secretory granules are large organelles, and not only are the mechanisms by which they fuse with the plasma membrane poorly known but also how cells maintain plasma membrane homeostasis after granule release. Biton *et al.* used advanced imaging techniques in *Drosophila* larval salivary glands to assess the machinery that governs the fusion pore dynamics in large exocrine secretory vesicles. Branched actin, myosin II, and BAR-domain proteins collaborate to control the fusion pore, facilitating a distinctive type of exocytosis called membrane crumpling that preserves apical membrane homeostasis during secretion. This mechanism allows large volumes of material to be secreted without disturbing the delicate balance of the cell's plasma membrane. —SMH

J. Cell Biol. (2023) 10.1083/jcb.202302112

Drosophila **salivary gland secretory granules fuse with the cell surface membrane and recruit actomyosin (magenta) to extrude the contents (green) by membrane crumpling.**

mechanisms underpinning ILC2s, but also reveal a previously unknown role for CD45 in the immune system. —SHR

Proc. Natl. Acad. Sci. U.S.A. (2023) 10.1073/pnas.2215941120

CATALYSIS

Cobalt's situation in hydroformylation

Cobalt carbonyl complexes are productive catalysts for the preparation of aldehydes by hydroformylation. However, they have been displaced by much more expensive rhodium in many applications because of their instability. A recent study reported that chelating bisphosphine ligands could greatly stabilize the cobalt catalysts, although a second study failed to reproduce those results and suggested instead that the phosphine-free carbonyl complex was more stable than previously thought. Hood *et al.* now revisit their original work with the chelating bisphosphine ligands and present multifaceted spectroscopic evidence that these are indeed active catalysts. They further propose that impure precursors may have hindered the prior reproduction effort. —JSY

J. Am. Chem. Soc. (2023) 10.1021/jacs.3c04866

PROTEIN SEQUENCING

Threading proteins through a nanopore

Nanopores are single-molecule biosensors and are widely used for ultrasensitive DNA sequencing. Uniform negative charges of single-stranded DNA allow an electric field to be used to move the DNA through the pore unidirectionally; however, proteins are chains of amino acids with heterogeneous charge distribution. Alternatively, osmotic pressure created by a net flux of water generated by a net ion charge on one side of the pore can be used to drive translocation of proteins. Sauciuc *et al.* show that such an electroosmotic flow approach can drive the transport of both unstructured model polypeptides and unfolded native proteins. —DJ

Nat. Biotechnol. (2023) 10.1038/s41587-023-01954-x

TIME-DOMAIN ASTRONOMY

A rapid luminous optical transient

The cadences of astronomical transient surveys enable the identification of unusual sources that evolve on time scales of a few days. Nicholl *et al.* report the detection of an optical transient that is as luminous as a superluminous supernova but evolves much more rapidly and has a different type of host galaxy. The source took 12 days to reach peak brightness and fade again afterward, which typically takes months for supernovae. The authors maintain that there are only two previously known transients with similar properties and that these three examples constitute a new class of sources. No theoretical models match all of their properties. —KTS

Astrophys. J. Lett. (2023) 10.3847/2041-8213/acf0ba

BIOSYNTHESIS

B_{12} bonanza

Vitamin B_{12} is an extremely complex molecule and performs some of the most fascinating chemistry in all of biology. It is an essential part of the animal food chain and is produced exclusively by microorganisms. Kang *et al.* developed a cell-free biosynthetic system to produce vitamin B_{12} consisting of four modules and a total of 26 individual enzymes. In addition to the biosynthetic enzymes, a complex regeneration scheme was used to replenish cofactors and cosubstrates. The amount of B_{12} produced in this way approaches what can be achieved with microbial fermentation, and this system should be amenable to further optimization and customization. —MAF

Nat. Commun. (2023) 10.1038/s41467-023-40932-4

RESEARCH ARTICLE SUMMARY

MICROBIOLOGY

PIM1 controls GBP1 activity to limit self-damage and to guard against pathogen infection

Daniel Fisch, Moritz M. Pfleiderer, Eleni Anastasakou, Gillian M. Mackie, Fabian Wendt, Xiangyang Liu, Barbara Clough, Samuel Lara-Reyna, Vesela Encheva, Ambrosius P. Snijders, Hironori Bando, Masahiro Yamamoto, Andrew D. Beggs, Jason Mercer, Avinash R. Shenoy, Bernd Wollscheid, Kendle M. Maslowski, Wojtek P. Galej, Eva-Maria Frickel*

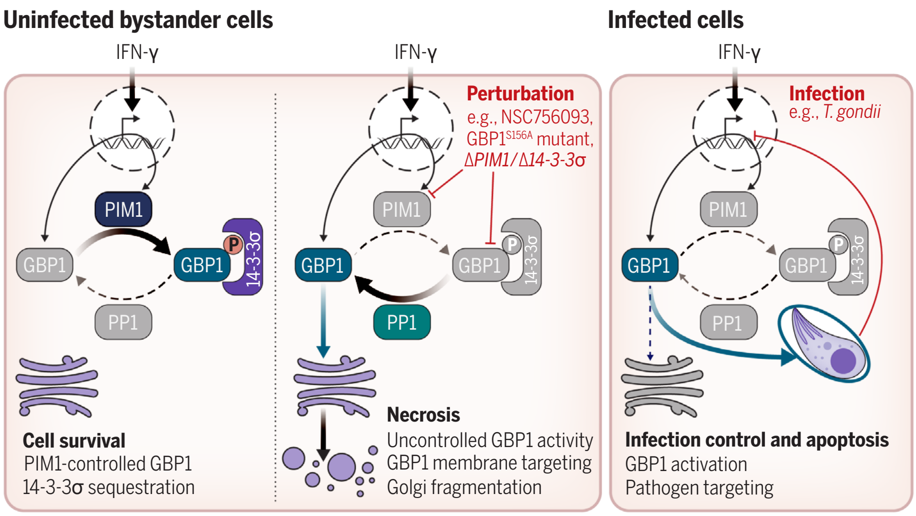
INTRODUCTION: Cells in infected tissues are exposed to inflammatory stimuli, including the innate and adaptive immunity–stimulating cytokine interferon-γ (IFN-γ). Although most tissue-resident and infiltrating cells are not infected, when exposed to IFN-γ, these bystander cells preemptively express a repertoire of interferon-stimulated genes (ISGs) with robust antimicrobial activities and the potential for self-harm. ISGs of the guanylate-binding protein (GBP) family are large, membrane-active guanosine triphosphatases (GTPases). GBPs can control intracellular microbes in various ways, most importantly by promoting membrane rupture and the release of microbial ligands and by the induction of programmed cell death, including pyroptosis and apoptosis. How uninfected cells protect themselves from the potentially self-destructive actions of GBPs while keeping these proteins readily available to combat infection is unknown.

RATIONALE: Cells need to tightly control the activity of antimicrobial proteins but rapidly deploy them upon infection. How is this achieved in human cells? Posttranslational modifications, such as phosphorylation by protein kinases, enable rapid and precise control of protein activities. We studied the phosphorylation of GBP1, a typical ISG, and how this modification affects its function, activity, and localization in human macrophages.

RESULTS: Ectopic expression of GBP1 in human macrophages led to changes in cell morphology, GBP1 accumulation at the Golgi apparatus, Golgi fragmentation, and uncontrolled cellular necrosis. These findings illustrate GBP1's potential to inflict self-damage. This phenotype was mitigated by IFN-γ treatment, suggesting that another IFN-γ–inducible factor limited GBP1 activity. We identified the kinase PIM1 as being this factor. We generated a phosphorylation-specific antibody and used high-resolution mass spectrometry to demonstrate GBP1 phosphorylation at serine-156 (Ser^{156}), which was guided by a basophilic PIM1 recognition motif. Ser^{156} is the central residue of a 14-3-3 protein binding motif, which suggests a switch-like function for its phosphorylation. Indeed, 14-3-3 proteins, especially 14-3-3σ, interacted with phosphorylated GBP1. In vitro reconstitution of this complex followed by single-particle cryo–electron microscopy confirmed a 14-3-3σ dimer grabbing onto the GBP1 GTPase domain. This binding locked GBP1 in a GTPase-inactive, monomeric state and restrained its activity in the macrophage cytosol. Expressing phosphorylation-deficient GBP1 mutants or mutants that could not be recognized by the kinase PIM1 or bound by 14-3-3σ led to uncontrolled GBP1 activation and subsequent cell death. Genetic depletion of either PIM1 or 14-3-3σ had similar outcomes, as did treatment with the GBP1:PIM1 interaction inhibitor NSC756093. Using the inhibitor in IFN-γ–activated patient-derived tumor organoids increased organoid death and prevented organoid reformation. Thus, we found that PIM1 and 14-3-3σ together controlled the activity of GBP1 in human cells. Disrupting PIM1-driven control of GBP1 has potential therapeutic implications for cancer therapy and innate immunity.

We observed that PIM1 mRNA and protein were extremely short-lived. Infection with the apicomplexan parasite *Toxoplasma gondii*, a pathogen that resides within intracellular vacuoles and blocks IFN-γ signaling by means of the effector protein TgIST, led to fast depletion of PIM1. This in turn reduced GBP1 Ser^{156} phosphorylation levels and liberated GBP1 from 14-3-3σ sequestration. High-throughput imaging revealed that GBP1 then rapidly targeted *Toxoplasma*-containing vacuoles to improve control of the infection.

CONCLUSION: The IFN-γ–induced, short-lived kinase PIM1 guards the integrity of IFN-γ signaling and protects self-membranes by regulating the activity of the potent antimicrobial effector GBP1. Pathogens that block IFN-γ signaling, thereby reducing the levels of PIM1, are then exposed to GBP1-driven innate immune control. The phosphoregulation of GBP1 by PIM1 reveals an IFN-γ–dependent control mechanism that protects uninfected bystander cells from self-inflicted innate immune damage during the process of pathogen elimination. ■



The protein kinase PIM1 controls GBP1 activity. In uninfected bystander cells, IFN-γ signaling induces expression of PIM1, which phosphorylates GBP1, subjecting it to sequestration by 14-3-3σ. This shunts aberrant GBP1 activity and ensures cell survival (left). Perturbation of this homeostasis leads to GBP1 dephosphorylation and Golgi targeting and disruption, culminating in necrosis (middle). Intracellular microbes that disrupt IFN-γ signaling, e.g., *T. gondii*, deplete PIM1 and liberate GBP1, which then attacks and controls the infection (right).

The list of author affiliations is available in the full article online.
*Corresponding author. Email: e.frickel@bham.ac.uk
Cite this article as D. Fisch *et al.*, *Science* **382**, eadg2253 (2023). DOI: 10.1126/science.adg2253

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adg2253

RESEARCH ARTICLE

MICROBIOLOGY

PIM1 controls GBP1 activity to limit self-damage and to guard against pathogen infection

Daniel Fisch[1,2], Moritz M. Pfleiderer[3], Eleni Anastasakou[3], Gillian M. Mackie[4], Fabian Wendt[5,6], Xiangyang Liu[3], Barbara Clough[2], Samuel Lara-Reyna[2], Vesela Encheva[7], Ambrosius P. Snijders[7,8], Hironori Bando[9,10], Masahiro Yamamoto[9,10], Andrew D. Beggs[11], Jason Mercer[2], Avinash R. Shenoy[12,13], Bernd Wollscheid[5,6], Kendle M. Maslowski[4,14,15,16], Wojtek P. Galej[3], Eva-Maria Frickel[1,2]*

Disruption of cellular activities by pathogen virulence factors can trigger innate immune responses. Interferon-γ (IFN-γ)–inducible antimicrobial factors, such as the guanylate binding proteins (GBPs), promote cell-intrinsic defense by attacking intracellular pathogens and by inducing programmed cell death. Working in human macrophages, we discovered that GBP1 expression in the absence of IFN-γ killed the cells and induced Golgi fragmentation. IFN-γ exposure improved macrophage survival through the activity of the kinase PIM1. PIM1 phosphorylated GBP1, leading to its sequestration by 14-3-3σ, which thereby prevented GBP1 membrane association. During *Toxoplasma gondii* infection, the virulence protein TgIST interfered with IFN-γ signaling and depleted PIM1, thereby increasing GBP1 activity. Although infected cells can restrain pathogens in a GBP1-dependent manner, this mechanism can protect uninfected bystander cells. Thus, PIM1 can provide a bait for pathogen virulence factors, guarding the integrity of IFN-γ signaling.

Interferon-γ (IFN-γ) is a pivotal cytokine that promotes innate and adaptive immunity during infection. IFN-γ acts on all nucleated cells to induce expression of a large repertoire of interferon-stimulated genes (ISGs) with potent antimicrobial activities (*1*). Pathogens are targeted by antimicrobial pathways after their detection by pattern recognition receptors (PRRs) (pattern-triggered immunity) or when they disrupt a regulator of an immune response (effector-triggered immunity). The latter mechanism, also referred to as guard immunity, was originally described in plants and suggests that innate immunity can indirectly detect virulence factors by their activities (*2*). The complexity of guard immunity in mammalian systems remains largely unexplored [reviewed in (*3*)]. Whether guard immunity exists in the context of inflammatory type II interferon production and signaling is unknown.

Among the ISGs are large guanosine triphosphatases (GTPases) of the GBP family (*4*). GBPs restrict infection by targeting intracellular microbes, thereby promoting their rupture and the release of microbial ligands; by redirecting autophagy and oxidative machinery to pathogen-containing compartments; and by removing replicative niches through induction of pyroptosis or apoptosis (*5*–*15*). IFN-γ up-regulates antimicrobial genes, including GBPs, in the primary infected cells and in neighboring, uninfected bystander cells. IFN-γ thus induces ubiquitous expression of GBPs in most cell types. Among GBPs, GBP1 is the most abundant in IFN-γ–primed cells. GBP1 has a C-terminal prenylation site that enables its membrane association (*16*) and forms homo- and heterodimers and membrane-bound coatomers with itself or with other GBPs (*17*, *18*). The GBP1 GTPase converts guanosine triphosphate (GTP) to guanosine monophosphate (GMP) in two steps (*19*). GBP1 can target membranes of the host Golgi apparatus (*16*, *20*) as well as plasma membrane–derived vacuoles containing *Toxoplasma gondii* (Tg) (*10*, *11*) or lipopolysaccharide (LPS) on cytosolic bacteria (*7*, *11*, *13*–*15*). However, how uninfected cells protect themselves from the potentially destructive actions of GBP1 remains unclear.

IFN-γ and phosphorylation control GBP1 activity and prevent cytotoxicity

To investigate how the activities of GBP1 are regulated in uninfected cells, we used a doxycycline (Dox)–inducible expression system in phorbol myristate acetate (PMA)–differentiated THP-1 macrophages (THP-1Δ*GBP1*+*GBP1*) (*11*). We first measured the survival of cells that ectopically expressed GBP1 in the absence of exposure to IFN-γ. Dox-induced ectopic expression of GBP1 was cytotoxic, progressively leading to ~75% cell loss over 7 days (Fig. 1A). Exposure of GBP1-expressing cells to IFN-γ reduced cell loss by ~55% (Fig. 1A). Treatment with IFN-γ in the absence of ectopic GBP1 expression resulted in minimal cell loss (Fig. 1A). Thus, we surmised that an unknown IFN-γ–dependent process and/or protein(s) must block GBP1-driven toxicity.

GBP1 contains three putative, surface-exposed phosphorylation sites: Ser^{156}, Ser^{157}, and Tyr^{427} (Fig. 1B). Immunoprecipitated Flag-GBP1 from macrophages probed for phosphorylated serine residues confirmed the presence of a phosphorylated GBP1 species, the amount of which increased upon exposure to IFN-γ (Fig. 1C). We hypothesized that IFN-γ–mediated phosphorylation of GBP1 could be a regulatory switch that suppresses the toxicity of GBP1. To test this possibility, we studied phosphorylation-deficient mutants of GBP1 [S156A (Ser^{156}→Ala), S157A (Ser^{157}→Ala), Y427F (Tyr^{427}→Phe)] expressed in Δ*GBP1* cells (fig. S1). Whereas S157A and Y427F had no effect on the dispersed, granular appearance of GBP1 within the cytosol, $GBP1^{S156A}$ accumulated in perinuclear areas (fig. S2A), which were revealed to be the Golgi apparatus by immunofluorescence assays (fig. S2B). Furthermore, $GBP1^{S156A}$-expressing cells were more elongated than cells expressing wild-type (WT) GBP1 (fig. S2C) and displayed changes in their cortical actin cytoskeleton (fig. S2D). Phosphorylation of GBP1 Ser^{156} might thus control localization and activities of GBP1.

To test whether GBP1 was phosphorylated at Ser^{156} specifically, we raised a phosphospecific antibody and validated its specificity with the following results: First, the antibody specifically interacted with phosphorylated GBP1 peptide (epitope: amino acids Ile^{152} to Cys^{166}); second, it bound to immunoprecipitated Flag-GBP1 WT but not Flag-$GBP1^{S156A}$; third, with the exception of minor reactivity toward GBP3, the antibody showed no cross-reactivity with other GBPs; and fourth, it specifically stained mCherry-tagged GBP1 (mCH-GBP1) but not mCH-$GBP1^{S156A}$ in immunofluorescence assays (fig. S3, A to D). Using this antibody, we confirmed that GBP1 was phosphorylated at Ser^{156} upon exposure of cells to IFN-γ by immunoprecipitation and immunoblot assays (Fig. 1D). Expression of $GBP1^{S156A}$ was more toxic than that of WT GBP1, and IFN-γ treatment no longer suppressed the toxicity of $GBP1^{S156A}$ (Fig. 1E). We thus checked

[1]Host-*Toxoplasma* Interaction Laboratory, The Francis Crick Institute, London, UK. [2]Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Edgbaston, UK. [3]European Molecular Biology Laboratory, 71 Avenue des Martyrs, Grenoble, France. [4]Institute of Immunology and Immunotherapy, University of Birmingham, Edgbaston, UK. [5]Department of Health Sciences and Technology (D-HEST), ETH Zürich, Institute of Translational Medicine (ITM), Zürich, Switzerland. [6]Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland. [7]Mass Spectrometry and Proteomics Platform, The Francis Crick Institute, London, UK. [8]Bruker Nederland BV, Leiderdorp, Netherlands. [9]Department of Immunoparasitology, Research Institute for Microbial Diseases, Osaka University, Osaka, Japan. [10]Laboratory of Immunoparasitology, WPI Immunology Frontier Research Center, Osaka University, Osaka, Japan. [11]Institute of Cancer and Genomic Sciences, University of Birmingham, Edgbaston, UK. [12]MRC Centre for Molecular Bacteriology and Infection, Department of Infectious Disease, Imperial College London, London, UK. [13]Inflammasome Biology Laboratory, The Francis Crick Institute, London, UK. [14]Institute of Metabolism and Systems Research, University of Birmingham, Edgbaston, UK. [15]Cancer Research UK Beatson Institute, Glasgow, UK. [16]School of Cancer Sciences, University of Glasgow, Glasgow, UK.
***Corresponding author. Email: e.frickel@bham.ac.uk**

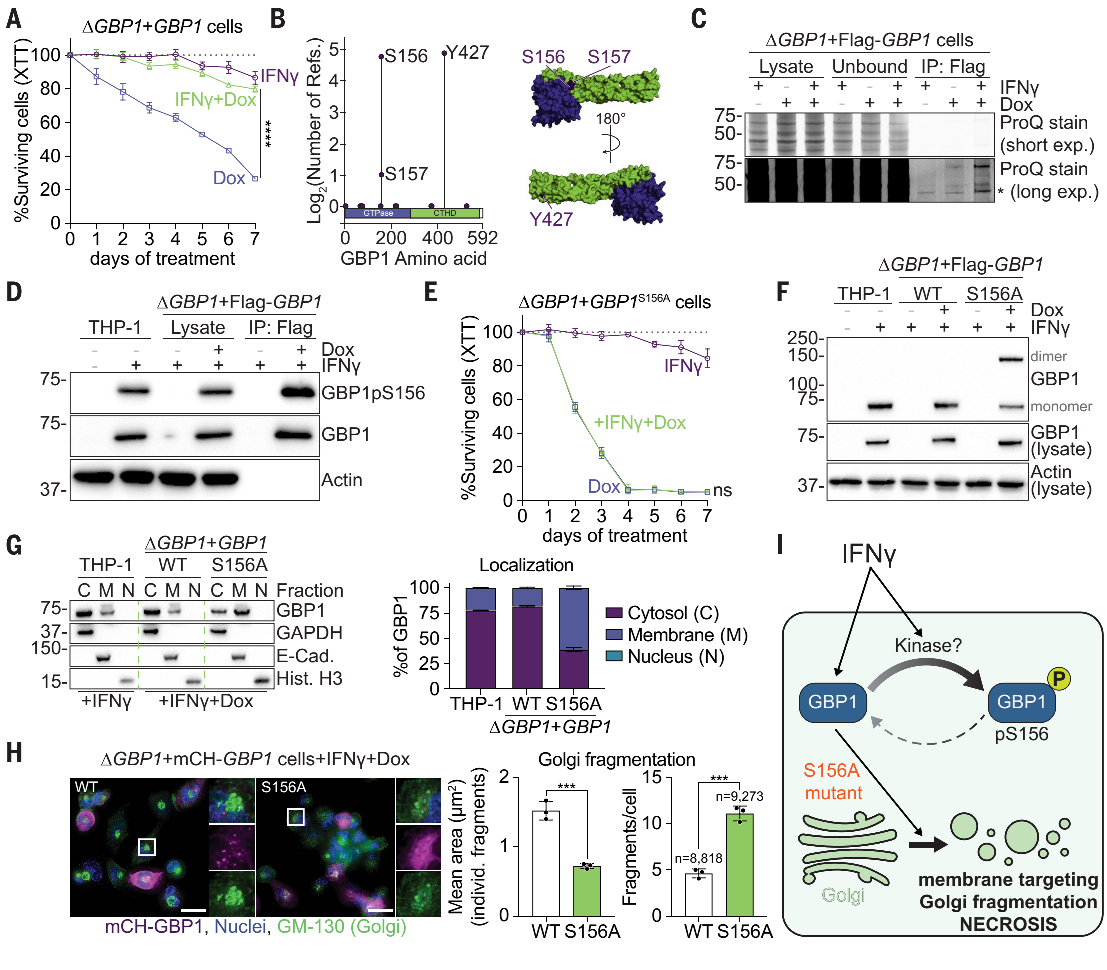


Fig. 1. IFN-γ and phosphorylation control GBP1 cytotoxicity and activity. (**A**) XTT cell survival assay of THP-1Δ*GBP1*+*GBP1* WT cells treated with IFN-γ with or without Dox for the indicated number of days. (**B**) Graph (left) and GBP1 structure (right) showing predicted phosphorylation sites and their surface-localization based on data from proteomics references (Refs). CTHD, C-terminal helical domain. (**C**) Images with short or long exposure (exp.) for a ProQ-Diamond phosphoprotein stain for immunoprecipitated Flag-GBP1 from THP-1Δ*GBP1*+Flag-*GBP1* cells treated with IFN-γ with or without Dox. The asterisk indicates background bands. IP, immunoprecipitation. (**D**) Immunoblots using GBP1pS156 antibody showing phosphorylation of GBP1 in IFN-γ– and Dox-treated THP-1Δ*GBP1*+Flag-*GBP1*. (**E**) XTT cell survival assay of THP-1Δ*GBP1*+*GBP1*^{S156A} cells treated with IFN-γ with or without Dox for the indicated number of days. (**F**) Immunoblot analysis of GBP1 oligomerization in cross-linked pellets and in cell lysates. (**G**) Immunoblots of subcellular fractionation to determine GBP1 localization in IFN-γ–primed THP-1 WT or Δ*GBP1*+*GBP1* cells expressing GBP1^{S156A} or WT GBP1 with markers for the cytosol (C), membranes (M), or the nucleus (N) (left) and quantification (right). E-Cad., E-cadherin; GAPDH, glyceraldehyde phosphate dehydrogenase. (**H**) Immunofluorescence images of cells expressing mCH-*GBP1* WT or mCH-*GBP1*^{S156A} for 48 hours and stained for Golgi marker GM130 (left) and quantification of mean area and number of Golgi fragments (right). The number of quantified cells is indicated in the figure. The smaller images on the right are zoomed-in views of the regions outlined by the white rectangles. Scale bars are 20 μm. Green, GM-130 (Golgi); magenta, mCH-GBP1; blue, nuclei. (**I**) Graphic summarizing findings on GBP1 activity control. Data information: Images in (C) and (D) and (F) to (H) are representative of n = 3 experiments. Graphs in (A), (E), (G), and (H) show mean ± SEM from n = 3 experiments. ***$P \leq 0.001$ and ****$P < 0.0001$ for the indicated comparisons in (A) and (E) from one-way analysis of variance (ANOVA) and in (H) from unpaired Student's t tests; ns is not significant. For gel source data, see data S5.

macrophages that expressed WT GBP1 or GBP1^{S156A} for markers of regulated cell death (fig. S4A) but did not observe any hallmarks of apoptosis, necroptosis, or pyroptosis after 2, 4, or 7 days (fig. S4B). Chemical inhibition of regulated cell-death pathways failed to prevent cells from dying over 4 days of constitutive ectopic expression of GBP1 (fig. S4C). Again, addition of IFN-γ rescued cells that expressed WT GBP1 but not GBP1^{S156A} and minimally affected cell viability by itself (fig. S4C). Thus, GBP1 phosphorylation on Ser^{156} was required to prevent unwanted toxicity and unregulated, necrotic death of human macrophages.

We next asked whether phosphorylation of GBP1 on Ser^{156} affects its known activities, such as oligomerization (*14*, *17*, *18*) and membrane association (*16*). Immunoblots showed that cytosolic WT GBP1 was monomeric, whereas GBP1^{S156A} dimers were readily detected, which is suggestive of increased GBP1 activity (Fig. 1F). Subcellular fractionation showed increased membrane association of GBP1^{S156A} (Fig. 1G). Because GBP1^{S156A} also exhibited increased Golgi targeting (fig. S2B), we next assessed Golgi morphology after 2 days of continuous GBP1 expression. We observed increased fragmentation of the Golgi in cells that expressed GBP1^{S156A} as compared with cells that expressed WT GBP1; the cells that expressed GBP1^{S156A} showed ~2.5 times more fragments per cell with a decreased average fragment size (Fig. 1H). Taken together, the phosphodeficient GBP1^{S156A} mutant showed increased association with and disruption of the Golgi, which was accompanied by unregulated cellular necrosis. Thus, Ser^{156} phosphorylation suppresses GBP1 activities and protects cells from self-inflicted damage (Fig. 1I).

The kinase PIM1 phosphorylates GBP1 at Ser^{156}

To narrow down the number of kinases that are capable of phosphorylating GBP1 on Ser^{156}, we first used in silico analysis, which identified nine candidate protein kinases (fig. S5A). For each of these kinases, we confirmed expression in macrophages by reverse transcription quantitative polymerase chain reaction (RT-qPCR) and reduced their expression with small interfering RNA (siRNA) silencing (figs. S5B and S6). Next, we immunoprecipitated Flag-GBP1 from THP-1 cells transfected with siRNAs directed against the kinases or against protein phosphatase 1/2 catalytic subunit alpha (*PPP1CA*, *PPP2CA*; Fig. 2A). Quantification of phosphoproteins suggested a reduction in

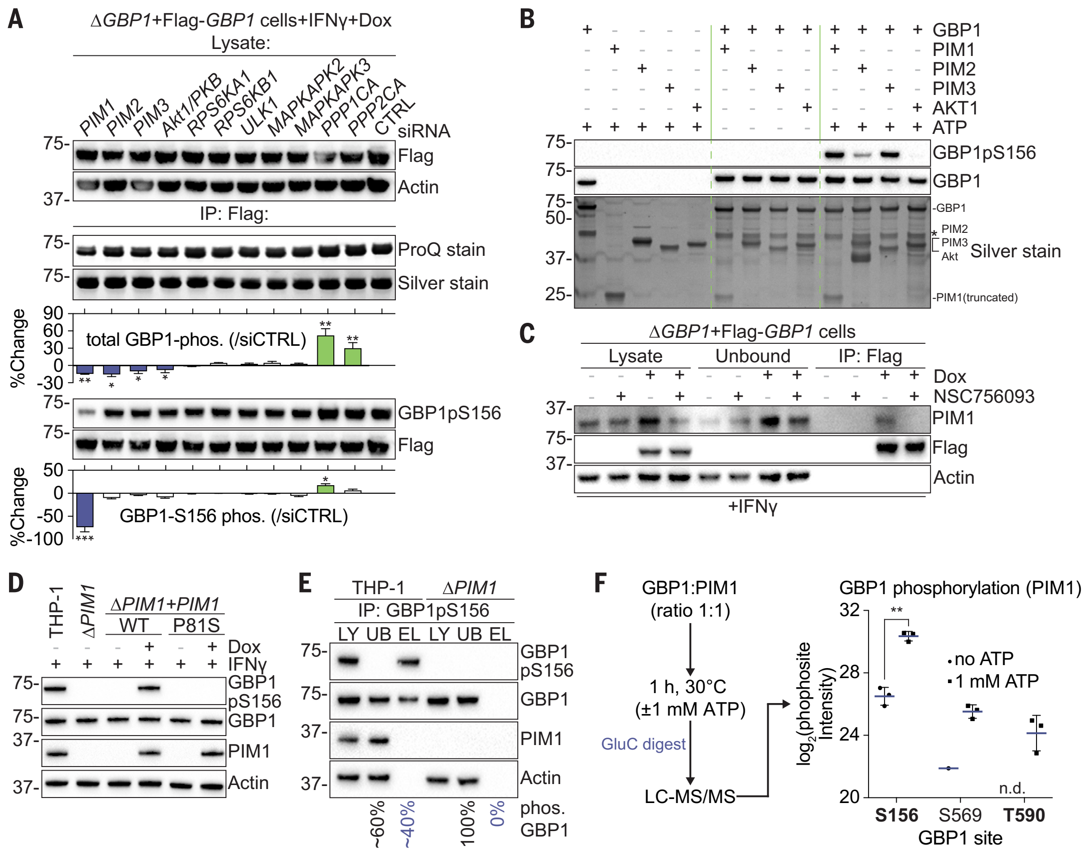


Fig. 2. The kinase PIM1 phosphorylates GBP1 at Ser^{156}. (**A**) Immunoblots and silver or ProQ-Diamond phosphoprotein stain of immunoprecipitated Flag-GBP1 WT from IFN-γ– and Dox-treated THP-1Δ*GBP1*+Flag-*GBP1* cells transfected with siRNA for the indicated genes. Quantification shows the change in total GBP1 phosphorylation (top) and the change in GBP1 Ser^{156} phosphorylation (bottom) normalized to siRNA control (siCTRL)–transfected cells. (**B**) Immunoblot and silver stain of in vitro kinase assays with PIM family kinases with or without 10 μM ATP. The asterisk indicates unspecific protein bands. (**C**) Immunoblots after cross-linking and immunoprecipitation of Flag-GBP1 from THP-1Δ*GBP1*+Flag-*GBP1* cells treated with IFN-γ with or without Dox and GBP1:PIM1 interaction inhibitor NSC756093. (**D**) Immunoblots testing for phosphorylation of endogenous GBP1 in IFN-γ–primed THP-1 WT or Δ*PIM1* macrophages reconstituted with PIM1 WT or a kinase-dead mutant (P81S) and additionally treated with Dox. (**E**) Immunoblots of GBP1 immunoprecipitation from IFN-γ–primed THP-1 WT or Δ*PIM1* cells using GBP1pS156 antibody. Volumes of unbound (UB) and elution (EL) fraction were equalized to determine the proportion of phosphorylated GBP1 (indicated below the immunoblot). LY, lysate. (**F**) MS analysis of the PIM1:GBP1 kinase assay. The plot depicts the $\log_2$ phosphosite intensity of identified Ser^{156}, Ser^{569}, and Thr^{590} sites of GBP1, which were localized with probabilities >0.99. LC-MS/MS, liquid chromatography–MS/MS; n.d., not detected. Data information: Images in (A) to (E) are representative of $n = 3$ experiments. Graphs in (A) show mean ± SD from $n = 3$ experiments, and graphs in (F) show mean ± SD from $n = 3$ replicates. $*P \le 0.05$, $**P \le 0.01$, $***P \le 0.001$, and $****P \le 0.0001$ in (A) from one-way ANOVA comparing all conditions with siRNA CTRL–transfected cells and in (F) from unpaired two sample t test after adjustment for multiple comparisons. For gel source data, see data S5.

total phosphorylation of GBP1 when PIM-family kinases were silenced, whereas silencing of *PPP1CA* or *PPP2CA* increased its total phosphorylation levels (Fig. 2A). Immunoblots using the Ser^{156} phosphospecific antibody showed that loss of PIM1 most strongly reduced phosphorylation of GBP1 at this residue (Fig. 2A). We next tested whether PIM family kinases could directly phosphorylate GBP1. Indeed, recombinant PIM1, 2, or 3 all phosphorylated GBP1 to varying degrees in vitro (Fig. 2B). Furthermore, we confirmed a previously observed direct interaction between PIM1 and GBP1 (*21*) in IFN-γ–exposed THP-1 cells after cross-linking and immunoprecipitation. The PIM1:GBP1 interaction inhibitor NSC756093 abolished this interaction (Fig. 2C). Next, we generated THP-1Δ*PIM1* CRISPR-Cas9 knockout cells and confirmed the knockout of PIM1 by immunoblotting and RT-qPCR (fig. S7A). Endogenous GBP1 was not phosphorylated at Ser^{156} in IFN-γ–exposed Δ*PIM1* cells, and phosphorylation was restored upon reconstitution with PIM1 WT but not with a PIM1 kinase-dead mutant [PIM1^{P81S} (*22*); Fig. 2D and fig. S7B]. Immunoprecipitation of endogenous GBP1 from IFN-γ–exposed THP-1 WT cells with the phosphospecific antibody showed that ~40% of the cellular GBP1 pool was phosphorylated (Fig. 2E).

PIM1, 2, and 3 are >65% similar to each other but vary in their efficiency of GBP1 phosphorylation (Fig. 2, A and B). To assess whether PIM2 and 3 also play a role in cellulo, we compared their expression levels in naïve and IFN-γ–treated macrophages by RT-qPCR and immunoblotting. Only PIM1 showed IFN-γ inducibility (fig. S8A), confirming previous findings (*23*). Knockout of PIM1 did not influence the kinetics or magnitude of GBP1 expression (fig. S8B), but endogenous GBP1 was no longer phosphorylated at Ser^{156} in Δ*PIM1* cells (Fig. 2D and fig. S8C). Additional depletion of PIM2 or PIM3 in Δ*PIM1* cells had no influence on phosphorylation of GBP1 as determined by immunoblotting (fig. S8C). However, immunoprecipitation and ProQ stain of endogenous GBP1 from Δ*PIM1* cells showed that the protein remained phosphorylated at residues other than Ser^{156} (fig. S8D).

Because PIM1 showed a prominent effect on the phosphorylation of GBP1 and is a known GBP1 interactor in cancer cells, we performed additional experiments with recombinant PIM1 and GBP1 to explore their interaction. These experiments showed a concentration-dependent increase in the phosphorylation of GBP1 by PIM1, which was abolished by inclusion of NSC756093, corroborating a requirement for direct interaction between the two proteins (fig. S8E). High-resolution mass spectrometry (MS) enabled dissection of the PIM1 kinase–GBP1 substrate relationship and identified three phosphorylation sites (Ser^{156}, Ser^{569}, and Thr^{590}; Fig. 2F). Fragment peak annotation of the electron-transfer dissociation (ETD)–activated tandem MS (MS/MS) spectra for the corresponding monophosphorylated peptides unequivocally confirmed phosphorylation at Ser^{156} and Thr^{590} (MaxQuant PTM site

localization probability >0.9999; fig. S8F and data S1). Ser[156] phosphorylation increased by 14-fold upon the addition of adenosine triphosphate (ATP), firmly establishing PIM1 as a bona fide GBP1 kinase.

We next sought to identify the PIM1 recognition sequence within GBP1. Evolutionary analysis of >3300 eukaryotic *GBP* genes identified the closest GBP1 homologs in 484 species (fig. S9A and data S2). Most of the analyzed organisms also had a well-conserved PIM1 homolog (data S2). The GBP1 homologs were analyzed for the presence of the corresponding phosphoserine, and the amino acid sequence upstream of the putative phosphoserine residue was compared with previously reported PIM1 recognition motifs (*24*–*27*). From this, it appeared that a basophilic kinase substrate motif ($R^{151}XRXKS^{156}$, where R is Arg, X is any residue, K is Lys, and S is Ser) was required for recognition of GBP1 by PIM1 (fig. S9A). The PIM1 motif and the putative phosphorylation site (residue equivalent to Ser[156]) were well conserved across evolution, and their appearance in GBP sequences correlated with each other (fig. S9A). We used a series of GBP1 mutants (fig. S1) to assess the functional role of this motif in guiding Ser[156] phosphorylation. We observed a complete loss of Ser[156] phosphorylation in K155A or R153A mutants and reduced phosphorylation for R151A (fig. S9B). Because Lys[155] and Arg[153] lie within the phosphospecific antibody epitope, their mutation could impair antibody binding. We thus used phosphoprotein-specific staining to assess phosphorylation independently. $GBP1^{K155A}$ had residual phosphoprotein-specific staining levels similar to that of $GBP1^{S156A}$, whereas $GBP1^{R153A}$ or $GBP1^{R151A}$ showed only a ~40% reduction (fig. S9B). Similarly, $GBP1^{R151A/R153A}$ and all mutants including K155A displayed reduced levels of phosphorylation, which were similar to the residual level of S156A (fig. S9B). Thus, the PIM1 phosphorylation motif in GBP1 is a basophilic kinase substrate motif.

Having established the GBP1 recognition motif for PIM1, we next assessed the other human GBP family members (fig. S9C). From this, it appeared that only GBP3, which is expressed at low levels in human macrophages (*11*), possesses a functional motif. All other GBPs have disruptive amino acid replacements in the recognition motif (GBP2, 4, 6, and 7) or a mutated phosphorylation acceptor site (GBP5; fig. S9C). We next tested the importance of PIM1 on cellular survival. In Δ*PIM1* cells, IFN-γ stimulation alone was sufficient to induce gradual cell loss over 7 days (fig. S9D), which resembled ectopic expression of WT GBP1 in naïve cells (Fig. 1A). Notably, the silencing of GBP1, but no other GBP, prevented viability loss, indicating that in the absence of PIM1, GBP1 was the only GBP activated in an uncontrolled manner (fig. S9D). IFN-γ–inducible PIM1 is thus directly responsible for phosphorylation of GBP1 on Ser[156], which prevents damage to cells.

14-3-3σ–bound phospho-GBP1 is inactive

Ser[156] of GBP1 is part of a flexible, surface-exposed loop and is the central serine of a putative, evolutionarily conserved mode I 14-3-3 binding motif (figs. S9A and S10A). MS analysis of GBP1-interacting proteins in IFN-γ–treated human macrophages showed the presence of all 14-3-3 proteins, including significant interactions with 14-3-3β, ζ, and σ (Fig. 3A and data S3), as expected (*28*). Immunoprecipitation of endogenous GBP1 confirmed its interaction with 14-3-3 proteins (pan–14-3-3 antibody) and with 14-3-3σ specifically (Fig. 3B). Immunoprecipitation of co-transfected Flag-GBP1 and hemagglutinin (HA)–tagged 14-3-3 proteins (β, γ, ε, ζ, η, θ, and σ) showed promiscuous interaction of all 14-3-3 proteins with GBP1 WT but not with $GBP1^{S156A}$ or $GBP1^{R153A/P158A}$ carrying a mutated 14-3-3 binding motif (fig. S10B). The $GBP1^{T590A}$ mutation did not affect its binding to 14-3-3 proteins (fig. S10B). We thus further examined the potential roles of Ser[156] phosphorylation. The mRNA levels for the seven 14-3-3 proteins varied in human macrophages, with the abundance of 14-3-3σ being the lowest in naïve cells and showing an increase upon IFN-γ treatment (figs. S6 and S10C). Expression of the other 14-3-3 family members remained unchanged (fig. S10C). Semiquantitative coimmunoprecipitations of phosphorylated endogenous GBP1 and purified HA-tagged 14-3-3 proteins corroborated our findings and showed the highest binding affinity between GBP1 and 14-3-3σ [dissociation constant (K_d) = 0.37 ± 0.10 μM (SD); fig. S11]. This was independently confirmed by isothermal titration calorimetry (ITC) with recombinant 14-3-3σ and in vitro phosphorylated recombinant GBP1 (Fig. 3C and fig. S12). Unphosphorylated GBP1 did not bind 14-3-3σ (Fig. 3C), confirming its specific interaction with the phosphorylated protein.

We next reconstituted the GBP1–PIM1–14-3-3σ reaction in vitro and found that these proteins formed a ~125-kDa complex in the presence of ATP, most likely composed of one monomer of phosphorylated GBP1 (67 kDa) and a 14-3-3σ dimer (dimer: 56 kDa) (fig. S13A). The stable phospho-GBP1:14-3-3σ complex could be purified and analyzed by size exclusion chromatography (fig. S13B). Dimerization of GBP1 occurred in absence of GTP but increased markedly in the presence of GTP or GDP plus AlF_3. The latter locks GBP1 in a GDP-bound state, showing that conformational changes in the GBP1 GTPase domain are required for its dimerization (fig. S13C). Phosphorylated and 14-3-3σ–bound GBP1 did not dimerize (fig. S13C). Furthermore, the rate of GTP hydrolysis by GBP1 in a GBP1:14-3-3σ complex was reduced by >8-fold as compared with hydrolysis by free GBP1 (Fig. 3D). Phosphorylation of GBP1 without 14-3-3σ binding did not influence the rate of GTP hydrolysis (Fig. 3D).

To visualize the interaction between phosphorylated GBP1 and 14-3-3σ, we solved the structure of the purified protein complex (fig. S13B) using single-particle cryo–electron microscopy (cryo-EM; Fig. 3E and fig. S14 and table S1 for details). We obtained a cryo-EM map of the GBP1:14-3-3σ heterotrimer at 5-Å resolution, which confirmed that one 14-3-3σ homodimer bound to a single copy of GBP1 by grabbing its GTPase domain and interacting with the linker helix that connects the helical domain to the GTPase domain (Fig. 3E). Although the overall resolution did not permit de novo modeling, we were able to unambiguously dock previously reported crystal structures as rigid bodies based on secondary-structure features (Fig. 3E). Close interactions were observed for GBP1 loop Arg[183] to Thr[197] with the termini of helices 8 and 9 of 14-3-3σ copy 1. Furthermore, 14-3-3σ helices 5 and 9 were in contact with the parallel-running linker helix of GBP1 (Fig. 3E). In this position, the phosphate binding site of 14-3-3σ copy 2 was in close proximity to GBP1 Thr[590], for which we also observed phosphorylation (Fig. 2F) but no effect on 14-3-3 binding upon mutation of GBP1 Thr[590] (fig. S10B). Thus, we hypothesize that 14-3-3σ binding to GBP1 functions in a similar way as 14-3-3 sequestration of Rnd3 GTPase (*29*), by requiring two phosphorylation sites and the farnesyl moiety. For GBP1, both Ser[156] and Thr[590] would be involved in sequential two-step 14-3-3-binding, with the initial recognition by Ser[156] and subsequent stabilization by binding to Thr[590]. In summary, our model suggests that phosphorylated GBP1 is sequestered by 14-3-3σ, which acts as a molecular padlock to keep the protein inactive through inhibition of dimerization and GTPase activity (Fig. 3F).

PIM1 and 14-3-3σ control GBP1 activities

Having identified a role for PIM1 and 14-3-3σ in inhibiting the activity of GBP1, we next asked whether PIM1 and 14-3-3σ protected cells from the toxic effects of GBP1. We used CRISPR-Cas9 to delete *14-3-3σ* (encoded by the gene *SFN*, also called *YWHAS*; fig. S7A) and reconstituted the cells with Dox-inducible 14-3-3σ (fig. S7C). Next, we measured the long-term survival of cells forced to express GBP1 mutants that cannot be recognized and inactivated by PIM1 ($GBP1^{R151A/R153A/K155A}$) or 14-3-3σ ($GBP1^{R153A/P158A}$). Similar to WT GBP1 (Fig. 1A), ectopic expression of these GBP1 mutants led to complete cell loss, pointing to their toxicity (Fig. 4A). Similar to phosphorylation-deficient $GBP1^{S156A}$ (Fig. 1E), IFN-γ treatment was unable to prevent cell death (Fig. 4A). We then measured the long-term survival of WT, Δ*GBP1*, Δ*PIM1*, Δ*PIM1/GBP1*, Δ*14-3-3*σ, and Δ*14-3-3σ/GBP1* cells. IFN-γ treatment

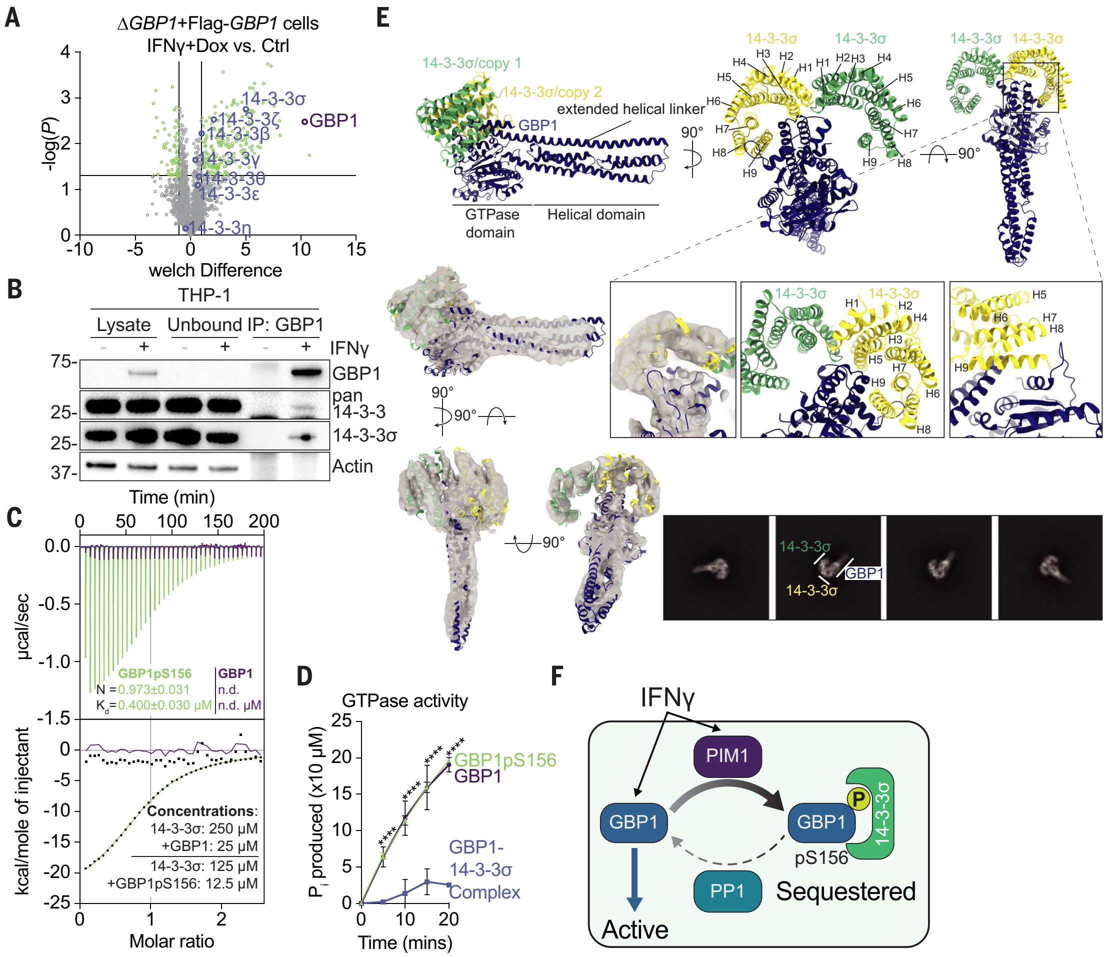


Fig. 3. Phosphorylated GBP1 is bound and inactivated by 14-3-3σ. (**A**) Volcano plot of MS data analysis of GBP1-interacting proteins obtained after coimmunoprecipitation of Flag-GBP1 from IFN-γ (Ctrl) or IFN-γ– and Dox-treated THP-1Δ*GBP1*+Flag-*GBP1* cells. GBP1-interacting proteins above the significance threshold are highlighted in green and 14-3-3 proteins in blue. (**B**) Immunoblots of coimmunoprecipitation of endogenous GBP1 from THP-1 WT cells treated with IFN-γ. (**C**) ITC determining thermodynamics of GBP1:14-3-3σ complex formation. 14-3-3σ at the indicated concentrations was injected in 7-μl aliquots to nonphosphorylated or in vitro phosphorylated GBP1 at the indicated concentrations. The determined molar ratio *N* at equilibrium and dissociation constant K_d are as indicated in the figure. n.d., not determined. (**D**) GTPase activity assay of 2 μM free GBP1, in vitro phosphorylated GBP1, or GBP1:14-3-3σ complex. P_i, inorganic phosphate. (**E**) (Top) Overview of the trimeric GBP1:14-3-3σ complex consisting of two 14-3-3 copies (yellow and green) bound to the GBP1 GTPase domain. (Middle) The insets depict details of the GBP1:14-3-3 dimer binding interface. The cryo-EM density is shown in gray with rigid-body docked GBP1 [Protein Data Bank (PDB) ID 1F5N] and 14-3-3σ (PDB ID 1YWT) crystal structures. (Bottom) Representative 2D classes with highlighted complex components. H, helix. (**F**) Graphic depicting the observed inhibitory mechanism, in which PIM1 phosphorylates human GBP1 at Ser[156], which is subsequently bound by 14-3-3σ. Upon dephosphorylation, GBP1 is liberated and becomes active. Data information: Images in (B) are representative of *n* = 3 experiments. Graphs in (A) and (D) show mean ± SD from *n* = 3 experiments. Curves in (C) are representative of *n* = 3 experiments and fitted to a model with one set of binding sites. ****$P < 0.0001$ for indicated comparisons in (D) from two-way ANOVA after adjustment for multiple comparisons. For gel source data, see data S5.

of Δ*PIM1* cells led to cell loss, as previously observed (fig. S9D), as did IFN-γ treatment of Δ*14-3-3σ* cells (Fig. 4B). In both cases, cell loss was prevented by additional knockout of GBP1, whereas IFN-γ treatment of THP-1 WT or Δ*GBP1* cells did not reduce viability (Fig. 4B). Moreover, both GBP1 mutants and endogenous GBP1 in Δ*PIM1* or Δ*14-3-3σ* cells were prone to dimerization (Fig. 4C) and localization to cellular membranes in IFN-γ–exposed macrophages (Fig. 4D). Again, this resembled the phosphorylation-deficient S156A mutant (Fig. 1, F and G).

To test whether this mechanism holds true at the tissue level, we used IFN-γ–responsive, patient-derived colorectal tumor organoids and tested whether the GBP1:PIM1 interaction inhibitor NSC756093 affected cell viability, organoid growth, and stemness. IFN-γ treatment alone had no impact on organoid survival or growth, and the inhibitor NSC756093 alone only caused mild cytotoxicity (Fig. 4E). However, treatment with a combination of IFN-γ and inhibitor (thereby unleashing GBP1 from its phosphorylation-driven control) decreased organoid growth by ~60% over the course of 4 days (Fig. 4E). We confirmed these findings by visual inspection, observing fewer and smaller organoids upon treatment with IFN-γ and NSC756093 compared with untreated or IFN-γ– or inhibitor-only treated samples. (Fig. 4F). To further interrogate the effects of treatment with IFN-γ and NSC756093, we next evaluated the stemness of organoids using stem-forming assays. To do so, organoids pretreated with IFN-γ and NSC756093 for 48 hours were dissociated, reseeded, and left to grow back into organoids for 7 days without any additional treatment (Fig. 4, G and H). This revealed a lasting effect of IFN-γ and NSC756093 on the ability of organoid precursors to self-renew and illustrates the potency of GBP1 in colorectal tumor organoids (Fig. 4, G and H). In summary, PIM1 and 14-3-3σ are required to control the activity of GBP1 in human cells by phosphorylation and cytosolic sequestration.

Toxoplasma infection depletes PIM1 and activates GBP1

We hypothesized that in the absence of phosphorylation of GBP1 by PIM1 and its sequestration by 14-3-3σ, GBP1 is free to target pathogen compartments more effectively. To test how control of GBP1 affects its trafficking and antimicrobial activities, we quantified its ability to restrict growth of Tg. We tested cells that expressed GBP1 mutants with impaired phosphorylation (S156A and R151A/R153A/K155A) or binding of 14-3-3σ (R153A/P158A), cells that lacked PIM1 (or expressing kinase-dead PIM1[P81S]), and

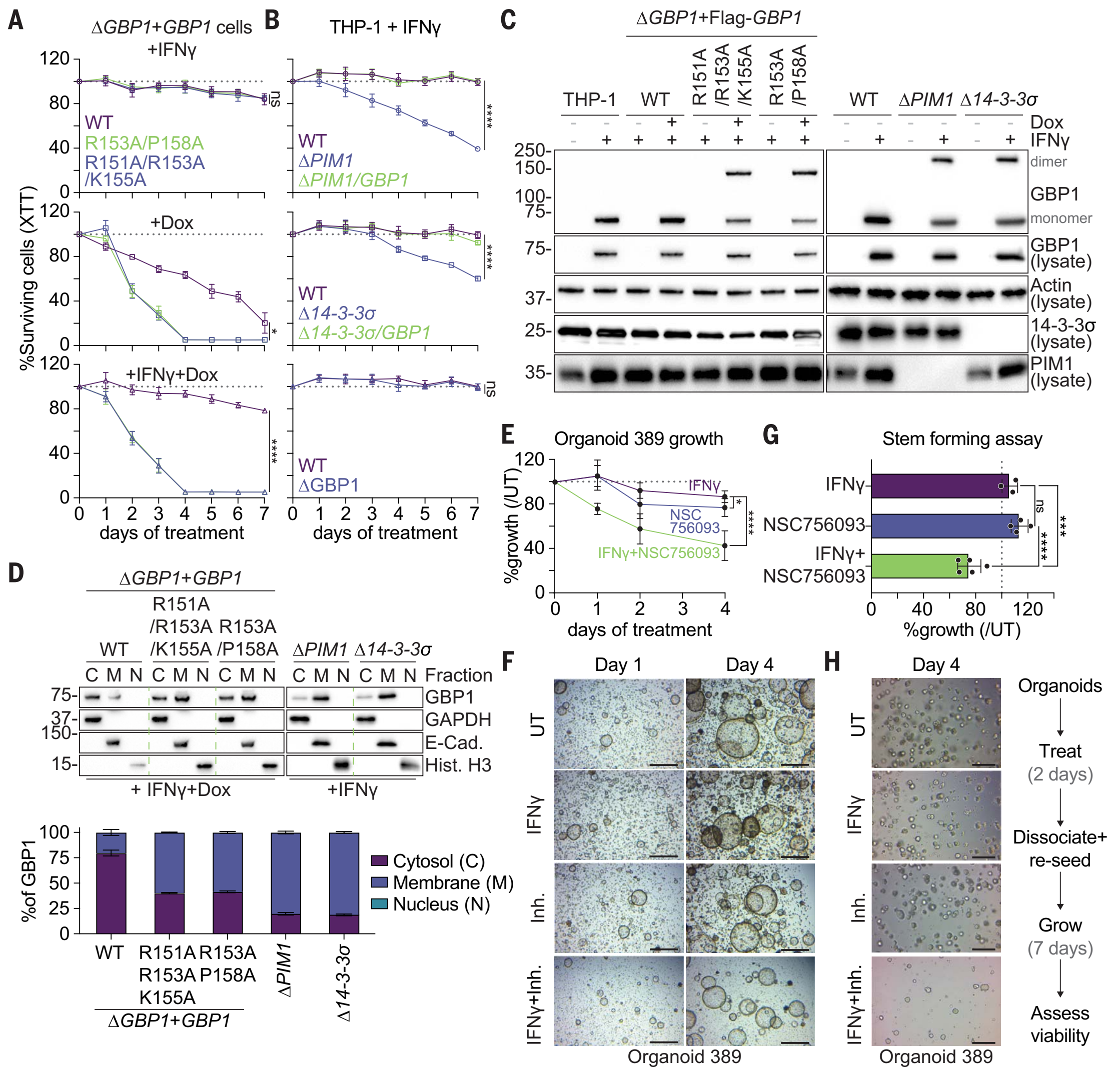


Fig. 4. PIM1 phosphorylation of GBP1 protects cells from self-inflicted damage. (**A**) XTT cell survival assay of THP-1Δ*GBP1*+*GBP1* WT, THP-1Δ*GBP1*+*GBP1*$^{R151A/R153A/K155A}$ (kinase motif mutant), or THP-1Δ*GBP1*+GBP1$^{R153A/P158A}$ (14-3-3 binding motif mutant) cells treated with IFN-γ with or without Dox for the indicated number of days. (**B**) XTT cell survival assay kinetics of THP-1Δ*GBP1*, Δ*PIM1*, Δ*14-3-3σ*, Δ*PIM1/GBP1*, or Δ*14-3-3σ/GBP1* cells treated with IFN-γ for the indicated number of days. (**C**) Immunoblot analysis of GBP1 in cross-linked pellets and in cell lysates of indicated cells treated with IFN-γ with or without Dox or left untreated. (**D**) Immunoblots of subcellular fractionation to determine GBP1 localization in indicated cells with respective markers for the cytosol (C), membranes (M), or the nucleus (N) (top) and quantification of localization (bottom). (**E**) MTT survival assay of patient-derived colorectal tumor organoids treated with IFN-γ and NSC756093 for 4 days, shown as percent growth compared with untreated (UT) controls. (**F**) Bright-field microscopy images of organoids after the indicated treatments with IFN-γ and inhibitor NSC756093 (Inh.). Scale bars are 1 mm. (**G**) Stem-forming assay of organoids treated for 2 days as indicated, dissociated, and regrown for 7 days in the absence of treatment. The graph shows the results of MTT assays on day 7 after reseeding. (**H**) Representative bright-field microscopy images from day 4 of regrowth of the stem-forming assay. Scale bars are 200 μm. Data information: Images in (C) and (D) and (F) to (H) are representative of $n = 3$ experiments. Graphs in (A), (B), (D), (E), and (G) show mean ± SEM from $n = 3$ experiments, normalized to untreated cells. $^*P \le 0.05$, $^{***}P \le 0.001$, and $^{****}P \le 0.0001$ in (A), (B), and (E) comparing all conditions with WT or untreated cells and for indicated comparisons in (E) and (G) from one-way ANOVA; ns is not significant. For gel source data, see data S5.

cells that lacked 14-3-3σ. All displayed increased recruitment of GBP1 to Tg vacuoles, restricted the growth of Tg more efficiently, and showed increased disruption of Tg-containing vacuoles and parasites, with a subsequent higher induction of apoptosis as compared with IFN-γ–treated THP-1 WT controls (Fig. 5A). These results indicate increased parasiticidal activity of GBP1 in conditions that prevent GBP1 phosphorylation and/or its interaction with 14-3-3σ. Reconstituting knockout cells with either WT PIM1 or 14-3-3σ decreased all pathogen control measures back to THP-1 WT levels (Fig. 5A).

When assessing the dynamics of the association of GBP1 with Tg-containing vacuoles, GBP1 was recruited faster and to higher levels in both Δ*PIM1* and NSC756093-treated THP-1 WT cells (fig. S15, A and B). Fluorescence recovery after photobleaching assays (FRAP) showed increased mobility of mCH-tagged GBP1 in Δ*PIM1* cells or cells that express mCH-GBP1^{S156A}, which could explain the faster kinetics of GBP1 recruitment to Tg vacuoles (fig. S15C). Conversely, the silencing of *PPP1CA* impaired the growth control of Tg, whereas treatment with NSC756093 improved it. Depletion of PIM2 or PIM3 affected neither GBP1 recruitment to Tg vacuoles nor growth control of Tg (fig. S15, D to F).

We next sought to determine how GBP1 phosphorylation and protein abundance changed in Tg-infected cells. We first measured the induction of GBP1, 14-3-3σ, and PIM1 as well as the appearance of Ser^{156}-phosphorylated GBP1 in THP-1 WT cells after IFN-γ treatment (Fig. 5B). This time course showed a more rapid induction of PIM1 than GBP1. Contrary to our RT-qPCR data (fig. S10C), these results showed that the levels of 14-3-3σ did not change (Fig. 5B). Notably, GBP1 appeared to be phosphorylated soon after it was expressed (Fig. 5B), indicating that the PIM1 and 14-3-3σ–mediated control of GBP1 was immediately active. PIM1 protein and mRNA were unstable and short-lived [half-life ($\tau_{1/2}$) ~8.5 or ~22 min, respectively] as compared with the more-stable GBP1 (protein $\tau_{1/2}$ > 12 hours; Fig. 5, C and D). This led us to hypothesize that PIM1 degradation might serve to uncouple the inhibition of GBP1. In this way,

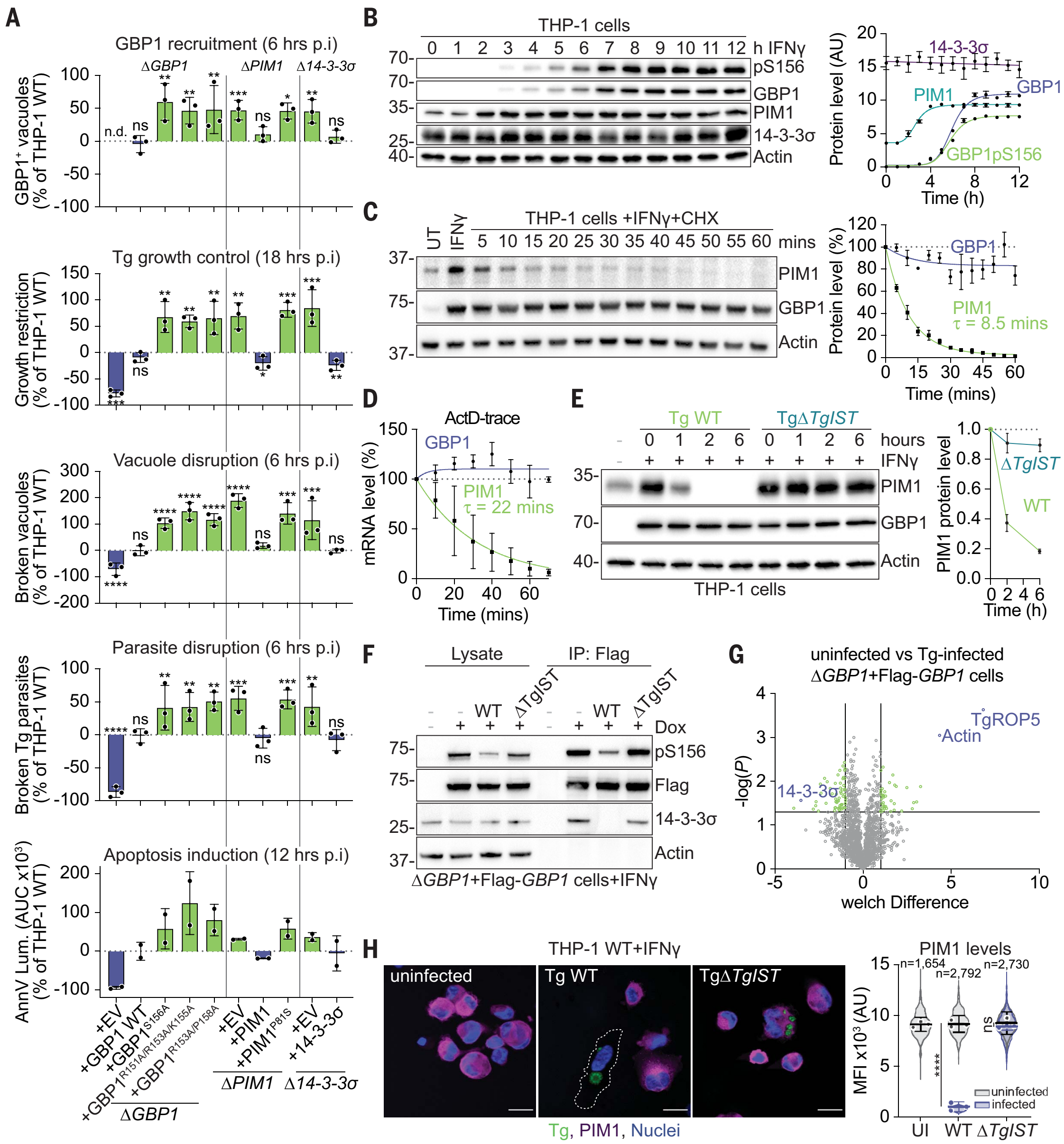


Fig. 5. *Toxoplasma* infection depletes PIM1 and activates GBP1. (**A**) Quantification of GBP1 recruitment to Tg vacuoles, Tg growth restriction, vacuole and parasite disruption, and apoptosis assay [area under the curve (AUC)] in IFN-γ– and Dox-treated cells expressing the indicated mutant of GBP1, Δ*PIM1* cells, or Δ*14-3-3σ* cells reconstituted with the indicated protein or empty vector (EV). Values plotted relative to IFN-γ–primed WT cells. n.d., not detected; p.i., postinfection. (**B**) Immunoblots of IFN-γ–treated THP-1 WT cells (left) and protein level quantification (right). (**C**) Immunoblots of THP-1 WT cells treated with 50 μg/ml cycloheximide (CHX) (left) and quantification of protein half-life (right). (**D**) RT-qPCR determination of mRNA stabilities in THP-1 WT cells treated with actinomycin D (ActD). (**E**) Immunoblots (left) and quantification (right) of PIM1 protein levels in THP-1 WT cells treated with IFN-γ and infected with Tg WT or Δ*TgIST*. (**F**) Immunoblots of Flag-GBP1 coimmunoprecipitation from IFN-γ– and Dox-treated THP-1Δ*GBP1*+Flag-*GBP1* cells infected with Tg WT or Δ*TgIST* for 6 hours. (**G**) MS analysis of GBP1-interacting proteins in uninfected or Tg WT–infected and IFN-γ– and Dox-treated THP-1Δ*GBP1*+Flag-*GBP1* cells. (**H**) Immunofluorescence images (left) of IFN-γ–primed and uninfected or Tg WT- or Δ*TgIST*-infected THP-1 WT cells after 12 hours. Scale bars are 20 μm. Dashed white line, cell outline; magenta, PIM1; green, Tg; blue, nuclei. The graph (right) shows quantification of PIM1 mean fluorescence intensity (MFI) depending on infection status. Gray, uninfected cells; blue, infected cells. Data information: Images in (B), (C), (E), (F), and (H) are representative of *n* = 3 experiments. Graphs in (A) show mean ± SD of *n* = 3 experiments. Graphs in (B) and (C) show mean ± SD from *n* = 3 experiments, and graphs in (D) and (E) show mean ± SD from *n* = 4 experiments. Graphs in (G) show data from *n* = 3 replicates. *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$, and ****$P \leq 0.0001$ in (A) from two-way ANOVA comparing all conditions with THP-1 WT cells and in (H) from nested *t* test comparing infected with uninfected cells after adjustment for multiple comparisons; ns is not significant. For gel source data, see data S5.

IFN-γ–induced PIM1 serves a dual role: It functions as a guard that detects pathogen-mediated interference of IFN-γ signaling (leading to depletion of PIM1), and it protects self-membranes from damage inflicted by GBP1. Tg secretes a pathogen effector protein called *T. gondii* inhibitor of STAT1 transcriptional activity (TgIST). Upon entering the host cell nucleus, TgIST sequesters STAT1 on specific loci and facilitates the formation of nonpermissive chromatin by recruiting the nucleosome remodeling deacetylase (NuRD) repressor complex. The nascent transcription sites thereby remain silenced (*30*, *31*). Given the IFN-γ–inducible nature of the *PIM1* gene, we hypothesized that TgIST completely shuts down *PIM1* gene expression, leading to the loss of PIM1. Indeed, we observed rapid TgIST-dependent depletion of PIM1 upon infection with Tg (Fig. 5E), further precipitating the loss of GBP1 phosphorylation and interaction with 14-3-3σ (Fig. 5F). MS of GBP1 interactomes of Tg-infected cells confirmed decreased interaction of GBP1 with 14-3-3σ and showed increased interaction with host-derived actin and parasite-derived TgROP5 compared with uninfected cells (Fig. 5G and data S4). These experiments show that the activity of GBP1 is restrained in uninfected cells but can be rapidly triggered upon infection by a pathogen that interferes with IFN-γ signaling. Supporting our hypothesis of bystander cell protection, we found that PIM1 was depleted specifically only in Tg WT–infected macrophages

but not in uninfected cells from that same experiment or in Δ*TgIST*-infected cells (Fig. 5H).

Finally, we wanted to investigate the effect of PIM1 and 14-3-3σ on GBP1 activity against Gram-negative bacteria such as *Salmonella* Typhimurium (STm), which is distinct from its activity against vacuolar pathogens such as Tg. For comparison, we assessed the cellular distribution of PIM1, 14-3-3σ, GBP1, and Ser^{156}-phosphorylated GBP1 in cells infected with Tg or STm by immunofluorescence imaging (fig. S16). In uninfected cells, all proteins showed a high degree of colocalization (Pearson's coefficient >0.88; fig. S16A). In Tg-infected cells, we observed the rapid recruitment of GBP1 to Tg vacuoles but no recruitment of PIM1, 14-3-3σ, or GBP1pS156 (GBP1 phosphorylated at Ser^{156}) (fig. S16B). This corroborated our finding that GBP1 must be dephosphorylated and released from sequestration before acquiring its parasiticidal, pathogen-proximal activity. We also observed decoration of the bacterial surface of cytosolic STm by GBP1 alongside recruitment of PIM1, 14-3-3σ, and GBP1pS156 (fig. S16C), suggesting that GBP1's high-affinity to LPS might override control by PIM1-mediated phosphorylation and 14-3-3σ sequestration. Supporting this finding, we did not observe degradation of PIM1 or dephosphorylation of GBP1 upon STm infection (fig. S16D). Finally, in Δ*14-3-3σ* cells, phosphorylated GBP1 was recruited to Tg vacuoles, demonstrating that retention by 14-3-3σ in the cytosol, not phosphorylation by PIM1 alone, ultimately controlled GBP1 pathogen-compartment targeting (fig. S16E).

Discussion

Human GBP1 is a membrane-active large GTPase that can disrupt microbe-encapsulating membranes (*5*–*15*). However, GBP1 can also associate with the Golgi apparatus (*16*, *20*) and/or the plasma membrane (*13*, *16*). How GBP1 activity is regulated in cells and how it is recruited to specific subcellular sites has remained elusive. In this work, we found that a large fraction of GBP1 is phosphorylated by the kinase PIM1 at Ser^{156}, establishing that posttranslational modification of human GBP1 regulates its activity. Ser^{156}-phosphorylated GBP1 was sequestered in the cytosol by 14-3-3σ, thereby preventing self-inflicted damage, in a mechanism reminiscent of how Rnd3 GTPase is inactivated (*29*). Thus, IFN-γ–exposed macrophages are in an activated state but keep GBP1 on standby and its activity tightly controlled. In uninfected cells, ectopic expression and unrestrained action of GBP1 led to necrosis. Similarly, ectopic expression of GBP1 in ovarian cancer cells reduces their viability by ~25% within 3 days (*32*). In infected cells, the Tg-induced block of IFN-γ signaling–induced transcription depletes inherently short-lived PIM1 (*33*) and springs a trap that unleashes GBP1 onto the pathogen-containing compartment. Pathogens from all kingdoms of life have evolved effector proteins that interfere with the integrity of IFN-γ signaling. Some examples include *Mycobacterium tuberculosis* blocking transcriptional transactivation downstream of STAT1 (*34*), *Vaccinia virus* phosphatase H1 targeting STAT1 for dephosphorylation (*35*), human cytomegalovirus depleting JAK kinases (*36*), and hepatitis C virus NS5A protein or *Leishmania donovani* both disrupting STAT1 phosphorylation (*37*, *38*). The cytokine-induced kinase PIM1 may thus serve as a detector for a broad range of pathogens that interfere with IFN-γ signaling. Consistent with our work, PIM1 is a host factor that favors survival of Tg (*39*, *40*). Furthermore, PIM1 is a known GBP1-interacting protein in cancer cells (*21*) and is a prosurvival proto-oncogene that negatively regulates other proapoptotic proteins (*24*, *26*, *41*–*44*). Given GBP1's ability to promote programmed cell death, use of the specific GBP1:PIM1 interaction inhibitor (*21*) is a promising way to chemically activate GBP1, with implications for innate immunity and cancer therapy.

Innate immune mechanisms that can cause cellular damage must be carefully regulated to ensure protection of bystander cells. The activity of immune effectors is tightly controlled, either indirectly through inactivation of inhibitory guard proteins or directly through activation of upstream PRRs. The indirect guard mechanisms, originally described in plants (*45*), detect the interference of pathogens with key cellular processes, whereas the direct PRR mechanisms respond to conserved pathogen-associated molecular patterns. An example of such "guarding" in Mammalia that is conceptually similar to the inactivation of GBP1 is the regulation of the Pyrin inflammasome, whereby the cytoskeleton-sensing kinases PKN1 and 2 phosphorylate Pyrin and inactivate the protein through the binding of 14-3-3 proteins. This system relays microbial perturbation of the cytoskeleton and thereby indirectly triggers immunity (*46*–*49*). Other recent examples include "self-guarding" of MORC3, which controls HSV-1 infection and secondary type I IFN responses (*50*), or viral disruption of protein synthesis that triggers apoptosis through the depletion of short-lived BCL-2 protein family members (*51*). Given the short half-life of PIM1, it is tempting to speculate that translation inhibition might also unleash GBP1 from its restriction. GBP1 directly activates immunity through its bona fide PRR function for cytosolic LPS (*7*, *11*, *13*–*15*). As we found in this work, its other, parasiticidal function is controlled by PIM1. Thus, GBP1 combines both innate immune surveillance strategies and antimicrobial activity in a single effector protein. The integration of an IFN-γ–induced pathway with posttranslational control of innate immune defense could cover a broad range of pathogens and enable cells to autonomously regulate immunity depending on their infection status. Similar cytokine-induced guard circuitries are likely to exist and thus ensure the protection of innocent bystander cells.

Materials and methods

Cell, parasite, and bacteria culture, infections, and treatments

THP-1 cells (TIB202, ATCC) were maintained in Roswell Park Memorial Institute (RPMI) medium with GlutaMAX (35050061, Gibco) supplemented with 10% fetal bovine serum (FBS) (Sigma), at 37°C in 5% CO_2. THP-1 cells were differentiated with 50 ng/ml phorbol 12-myristate 13-acetate (PMA) (P1585, Sigma) for 3 days and left to rest for 2 days by replacing the differentiation medium with complete medium without PMA. Human embryonic kidney 293T (HEK293T) cells (Cell Services, The Francis Crick Institute, London, UK), NIH3T3 cells (CRL-1658, ATCC), and human foreskin fibroblasts (HFFs) (SCRC-1041, ATCC) were cultured in Dulbecco's modified Eagle's medium (DMEM) with GlutaMAX (Gibco) supplemented with 10% FBS (Sigma) at 37°C in 5% CO_2. Cells were not used beyond passage 25 and HFFs not beyond passage 15.

Patient-derived colorectal tumor organoids (Organoid 389), originally derived from a T3N1 poorly differentiated, mismatch repair–proficient adenocarcinoma of the sigmoid colon from a male subject (*52*), were grown and maintained in 50 µl of Matrigel domes (CLS356231, Corning) with 500 µl of human IntestiCult medium (06010, StemCell) in 24-well plates at 37°C and 5% CO_2 and passaged weekly.

Tg were maintained by serial passage on HFF cells that were cultured in DMEM with GlutaMAX supplemented with 10% FBS at 37°C in 5% CO_2. Parasites were passaged the day before infection to maintain high viability. Tg were prepared from freshly 25-gauge-syringe–lysed HFF cultures by centrifugation at 50*g* for 3 min; the cleared supernatant was transferred into a new tube and followed by subsequent centrifugation at 500*g* for 7 min and resuspension of the pelleted parasites into fresh complete medium. Parasites were added to the cells at a multiplicity of infection (MOI) of 1. Infection was synchronized by centrifugation at 500*g* for 5 min. Two hours after infection, extracellular parasites were removed with three washes using warm phosphate-buffered saline (PBS) (806552, Sigma), and fresh complete medium was added before culturing at 37°C and 5% CO_2 for the required time.

STm strain SL1344-GFP (GFP, green fluorescent protein) (*11*) was maintained under ampicillin selection (11593027, Gibco) and grown on LB plus ampicillin agar plates. One day before infection, bacteria from a single colony were inoculated and grown overnight at 37°C. The overnight culture was diluted 1:20 into LB plus 300 mM NaCl and grown with shaking in a closed container to an optical density at 600 nm (OD_{600}) of 0.9. Bacteria were harvested

by centrifugation at 1000*g* for 5 min, washed with serum-free cell culture medium twice, and resuspended in 1 ml of plain medium. Cells were infected with STm at an MOI of 10, and infections were synchronized by centrifugation at 750*g* for 10 min. Infected cells were washed three times with warm PBS 30 min after infection, and fresh medium containing 10 μg/ml gentamicin (15750060, Gibco) was added. The bacterial MOI used for infections was confirmed by plating on LB agar plates. An overview of all cell lines, parasites, and bacteria strains can be found in table S2. All cell culture work was performed without the addition of antibiotics, and the cells were regularly tested for mycoplasma contamination by immunofluorescence and PCR.

THP-1 cells were stimulated for 16 hours before infection in complete medium with addition of 50 international units (IU)/ml human IFN-γ (285-IF, R&D Systems). Induction of protein expression in the Dox-inducible cells was performed with 200 ng/ml Dox (D9891, Sigma) overnight. To block translation for studying protein stability, cells were treated with 50 μg/ml cycloheximide (CHX) (C7698, Sigma), and to block transcription, cells were treated with 6 ng/ml actinomycin D (A1410, Sigma). To block the GBP1:PIM1 interaction, cells were treated with 100 nM NSC756093 (SML1310, Sigma). For in vitro assays, NSC756093 was used at 1 μM. To chemically induce pyroptosis, cells were primed with 100 ng/ml LPS O111:B4 (NC9673121, Enzo) and subsequently treated with 10 μM nigericin. Treatments to induce apoptosis were performed with 50 ng/ml tumor necrosis factor–α (TNFα) (210-TA, R&D Systems) and 10 μg/ml CHX. The pan-caspase inhibitors zVAD-fmk (25 μM, 60332, Cell Signaling Technologies) or qVD (20 μM, ab141421, Abcam) were used to block caspase activation for necroptosis induction or blocking apoptosis or pyroptosis. To inhibit pyroptosis, cells were treated with 10 μM Mcc950 (inh-mcc, InvivoGen) or 50 μM disulfiram (DSF) (1224008, Sigma). To inhibit necroptosis, cells were treated with 50 μM GSK'872 (6492, Tocris) or 10 μM necrosulfonamide (NSA) (5025, Tocris). Tumor organoid cultures were treated with 100 IU/ml IFN-γ or 1.66 μM NSC756093.

Plasmid DNA and siRNA transfection

THP-1 cells were transfected with 30 nM siRNAs 2 days before infection (table S3). The transfection mix was prepared as a 10× mix in OptiMEM containing the siRNA(s) and TransIT-X2 transfection reagent (MIR600x, Mirus) in a 1:2 ratio. All siRNAs were ON-TARGETplus pools from Dharmacon, with the negative control being ON-TARGETplus Non-targeting Pool (D-001810, Dharmacon). Transfection of DNA plasmids was performed using Lipofectamine 2000, following the manufacturer's instructions (11668027, Invitrogen).

Lentiviral transduction

Lentiviral packaging used HEK293T cells transfected with equimolar ratios of the required expression plasmid, pMD2.G, and psPAX2 (Addgene, 12259 and 12260, respectively, both gifts from D. Trono) using Lipofectamine 2000 in serum-free DMEM. Medium was replaced 12 hours after transfection with fresh DMEM plus 10% FBS to rest the cells and again after 12 hours with DMEM plus 10% FBS containing 5 mM sodium butyrate (B5887, Sigma). The next day, medium was replaced with RPMI medium plus 10% FBS, and the cells were left to produce lentiviral particles for 1 day. Virus containing supernatant was filtered through a 0.43-μm syringe filter and supplemented with 8 μM polybrene (H9268, Sigma). The target cells were then resuspended in 500 μl of the virus-containing medium and "spinfected" for 30 min at 1000*g*. One hour after infection, 1 ml of complete medium was added, and the cells were left to rest. The infection procedure was repeated a total of three times before selection with the appropriate selection reagent: 1 μg/ml puromycin (A1113802, Gibco), 200 μg/ml zeocin (J67140, Alfa Aesar), and/or 15 μg/ml blasticidin S (15205, Sigma). Once untransduced control cells had died in the selection medium, the newly created cells were verified for successful transduction by immunoblotting.

Creation of new cell lines

Inducible GBP1 cell lines

THP-1Δ*GBP1*, THP-1Δ*GBP1*+Tet-*GBP1* WT, THP-1Δ*GBP1*+Tet-Flag-*GBP1* WT, and THP-1Δ*GBP1*+Tet-mCH-*GBP1* WT have been published before, and new cell lines were created identically using lentiviral transductions (*11*). To make cells expressing mutated GBP1, pLenti-Tet-GBP1 plasmids (*11*), with and without tags, were mutated using site-directed mutagenesis and transduced into the THP-1Δ*GBP1*+Tet target cells using lentiviral transduction as described before.

PIM1 and 14-3-3σ (SFN/YWHAS) CRISPR knockout cells

Guide RNA (gRNA) sequences targeting the 5′ and 3′ untranslated regions (UTRs) of the respective genes were designed using the MIT CRISPR design tool (crispr.mit.edu) (table S4). DNA oligonucleotides encoding the CRISPR RNA (crRNA) were annealed and cloned into BsmBI-digested (ER0451, Thermo Fisher Scientific) pLentiCRISPR-V2 backbone (*53*) using Quick Ligation kit (M2200, New England Biolabs) and transduced into THP-1 WT or THP-1Δ*GBP1* cells using lentiviral particles. After selection with puromycin for 7 days, cells were subcloned by serial dilution into 96-well plates using preconditioned complete medium supplemented with nonessential amino acids (11140076, Gibco), penicillin and streptomycin, and GlutaMAX. Roughly 3 weeks after seeding, obtained clones were expanded and screened for the absence of target protein expression by RT-qPCR. Clones that showed reduced or absent target protein expression underwent secondary screening by immunoblotting. For each of the cell lines, >4 single-cell clones with confirmed absence of either PIM1 or 14-3-3σ protein were pooled and cultured for 2 more weeks before undergoing final screening to confirm absence of the proteins. This resulted in the THP-1Δ*PIM1*, Δ*PIM1*/*GBP1*, Δ*14-3-3*σ, and Δ*14-3-3*σ/*GBP1* cell lines.

Reconstitution of PIM1 and 14-3-3σ

CRISPR knockout cells of PIM1 and 14-3-3σ were transduced with the Dox-inducible system as previously described (*11*). To create PIM1- and 14-3-3σ–expressing Dox-inducible plasmids (pLenti-Tet-*PIM1* or pLenti-Tet-*14-3-3*σ), the empty vector backbone was digested with BamHI, PIM1 and 14-3-3σ open reading frames (ORFs) were amplified from cDNA obtained from IFN-γ–primed THP-1 WT cells, and all fragments assembled with Gibson assembly. Further, the pLenti-Tet-PIM1 plasmid was mutated by site-directed mutagenesis to obtain a pLenti-Tet-$PIM1^{P81S}$ kinase-dead version, and all vectors transduced into the respective target cells using lentiviral transduction.

RT-qPCR

RNA was extracted from 0.25×10^6 cells using Trizol reagent (15596026, Invitrogen). GlycoBlue (5 μg/ml, AM9516, Invitrogen) was added during the isopropanol precipitation step. RNA quality was measured on a Nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific). One microgram of RNA was reverse transcribed using a high-capacity cDNA synthesis kit (4368813, Applied Biosystems). For qPCR, the PowerUP SYBR green kit (A25742, Applied Biosystems), 20 ng cDNA in a 10-μl reaction, and primers at 1 μM final concentration (sequences in table S5) were used. Primer specificity was ensured by designing primers to span exon-exon junctions, whenever possible, and for each primer pair, a melt curve was recorded and compared with in silico predicted melt curves from uMelt (*54*) and amplicon sizes were analyzed by agarose gel electrophoresis. Recorded cycle threshold (Ct) values were normalized to the recorded Ct of human *HPRT1*, and data were plotted as ΔCt (relative expression). To determine absolute expression, defined amounts of linearized plasmid standards were used as PCR template and the obtained Ct values used to calculate transcript numbers.

MTT, XTT, and AnnV glow assays for cell survival and death

To determine cell viability using XTT assays (Cell Proliferation Kit II, Roche), 25,000 cells were seeded per well in a 96-well plate or 10,000 cells in a 384-well plate and differentiated as described before. Cells were treated in RPMI

without phenol red (11835, Gibco). The detection reagent was freshly prepared according to the manufacturer's instructions and added before the viability determination. Cells were then incubated at 37°C for 4 hours, and absorption was measured at 475 and 660 nm for correction.

For organoid growth analysis, MTT reagent (M6494, Invitrogen) was added in a 1:10 dilution to the plates, and the Matrigel domes were disrupted by pipetting. Plates were returned to a 37°C incubator for 30 min before adding lysis solution (4 mM HCl, 0.1% Triton-X100 in isopropanol), mixing by pipetting, and placing on a shaker in the dark for 10 min. Each well was thoroughly mixed, lysates were transferred in quadruplicates to a microplate, and absorbance was read at 550 nm on an iMark Microplate Absorbance Reader (Bio-Rad). Medium with Matrigel was used as a negative control.

Apoptosis kinetics were analyzed using the RealTime-Glo Annexin V Apoptosis Assay (JA1001, Promega) according to the manufacturer's instructions. In brief, 50,000 cells were seeded per well of a white, tissue culture–treated 96-well plate, differentiated, pretreated, and infected. Simultaneously with infection, detection reagent was added. Luminescence was measured using a Fluostar Omega plate reader (BMG Labtech) that was preheated to 37°C. No-cell, medium-only controls were used for background correction. To determine overall cell death levels, the area under the curve (AUC) was determined using Prism 8.4 (GraphPad, Inc.).

Molecular cloning of vectors for transient expression

For transient expression of Flag-*GBP1* after transfection, the ORF was amplified by PCR using Q5 polymerase and cloned into BamHI- and EcoRI-digested pcDNA3.1(+) vector using the Quick Ligation kit. pcDNA3-HA-14-3-3β, γ, ε, ζ, and σ were a gift from A. Ridley (University of Bristol, Bristol, UK) (*29*) and M. Yaffe (MIT, Boston, MA, USA) (*55*). To produce N-terminally HA-tagged 14-3-3η and 14-3-3θ in a pcDNA3 vector identical to the other constructs, the pcDNA3-HA backbone was amplified by PCR. 14-3-3η was amplified from pCS2-HA-14-3-3η (Addgene, 116887, a gift from F.-Q. Li and K.-I. Takemaru) (*56*) and 14-3-3θ from pcDNA3-14-3-3θ-HA (a gift from A. Ridley), and finally, inserts and backbone were assembled by Gibson assembly.

SDS-PAGE, immunoblotting, and gel staining

For immunoblotting, 0.5×10^6 cells were seeded per well of a 48-well plate, differentiated with PMA, pretreated, and infected as described above. At the end of treatments, cells were washed with ice-cold PBS and lysed for 5 min on ice in 50 μl of RIPA buffer (150 mM NaCl, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.1% SDS, 25 mM Tris-HCl pH 7.4) supplemented with protease inhibitors (Protease Inhibitor Cocktail set III, EDTA free, Merck) and PhosSTOP phosphatase inhibitors (4906845001, Roche). Lysates were cleared by centrifugation at full speed for 15 min at 4°C. Subcellular fractionation was performed using the QProteome Cell Compartment Kit (37502, Qiagen) following the manufacturer's instructions. Bicinchoninic acid (BCA) assay (Pierce BCA protein assay kit, 23225, Thermo Fisher Scientific) was performed to determine protein concentrations. Ten micrograms of total protein per sample were mixed with Laemmli buffer (1610737, BioRad) containing 5% dithiothreitol (DTT) and denatured at 95°C for 10 min and then run on Bis-Tris gels (Novex, Invitrogen) in MOPS running buffer. For immunoblots of culture supernatants, cells were treated in OptiMEM (1105802, Gibco). Proteins in the supernatants were precipitated with 4 Vol acetone (V800023, Sigma) overnight at –20°C and pelleted by centrifugation. Pellets were air dried for 10 min, resuspended in 40 μl of 2× Laemmli loading dye, denatured, and used for immunoblotting.

After SDS–polyacrylamide gel electrophoresis (SDS-PAGE), proteins were transferred onto nitrocellulose membranes using the iBlot transfer system (Invitrogen). Depending on the primary antibody used, the membranes were blocked with either 5% bovine serum albumin (BSA) (A2058, Sigma) or 5% dry-milk (M7409, Sigma) in TBS-T (0.05% Tween-20) for at least 1 hour at room temperature. Incubation with primary antibodies (table S6) was performed at 4°C overnight. Blots were developed by washing with TBS-T, probed with 1:5000 diluted HRP-conjugated secondary antibodies, washed again and imaged on a ChemiDoc MP imaging system (BioRad) using enhanced chemiluminescence (ECL) (Immobilon Western, WBKLS0500, Millipore). For quantification of protein band intensities, images were quantified using FIJI/ImageJ and normalized to the actin loading control of each membrane.

For silver staining of protein gels, after SDS-PAGE, the gels were washed in ddH_2O and then silver stained following the manufacturer's instructions (Silver Stain Plus Kit, 1610449, BioRad). Similarly, washed gels were used for ProQ Diamond staining following the manufacturer's instructions (Pro-Q Diamond Phosphoprotein Gel Stain, P33300, Invitrogen). Stained gels were imaged on a ChemiDoc MP imaging system (BioRad). Similar to chemiluminescence quantification, protein bands from stained gels were quantified using FIJI ImageJ. Coomassie stain of SDS-PAGE gels was performed using InstantBlue Coomassie Protein Stain (ISB1L, Abcam).

GBP1 oligomerization assay

To assess GBP1 oligomerization, a protocol previously published for ASC oligomerization (*57*) was adapted. In brief, cells were seeded and treated as described above but harvested by scraping in ice-cold PBS containing 2 mM EDTA. Next, cell pellets were washed with PBS plus 2 mM EDTA and resuspended in 200 μl of buffer A (20 mM HEPES-KOH pH 7.5, 10 mM KCl, 1.5 mM $MgCl_2$, 1 mM EDTA, 1 mM EGTA, 320 mM sucrose) and lysed by passing the suspension through a 25-gauge needle. Lysates were cleared by centrifugation at 1800*g* and 4°C for 15 min. From this, sample was kept as lysate input control. The remaining supernatant was diluted 2× with buffer A, and then 400 μl of CHAPS buffer (20 mM HEPES-KOH pH 7.5, 5 mM $MgCl_2$, 0.5 mM EGTA, 0.1% CHAPS) was added. Precipitated GBP1 was pelleted by centrifugation at 7500*g* for 30 min. Finally, pellets were resuspended in 30 μl of CHAPS buffer with 4 mM of disuccinimidyl suberate (DSS) (21655, Thermo Fisher Scientific) and left to cross-link for 30 min at room temperature, pelleted again by centrifugation at 7500*g* for 30 min and then dissolved in 2× Laemmli loading dye. Samples were then analyzed by immunoblotting.

Immunoprecipitations

For immunoprecipitations 5×10^6 cells were seeded in six-well plates and differentiated, pretreated, and infected as described above. The cells were washed in ice-cold PBS and scraped from the plates. Whole-cell lysates were prepared by adding 500 μl of lysis buffer (1% Triton X-100, 20 mM Tris–HCl pH 8.0, 130 mM NaCl, 1 mM DTT, 10 mM sodium fluoride) with added protease and phosphatase inhibitor cocktails and incubating on ice for 15 min. Lysates were cleared by centrifugation at 4°C for 10 min at full speed.

For Flag and HA immunoprecipitations, cleared lysates were added to Flag(M2)-agarose beads (A2220, Sigma) or Pierce Anti-HA Agarose beads (26182, Thermo Fisher Scientific) pre-equilibrated by washing three times with lysis buffer. Proteins were captured by incubation on a rotator overnight at 4°C. Beads were washed once with lysis buffer, three times with lysis buffer containing 260 mM NaCl, and again twice with lysis buffer. Proteins were eluted using 200 ng/ml 3xFlag peptide (F4799, Sigma) or 200 ng/ml HA peptide (I2149, Sigma) in lysis buffer by incubation on an orbital shaker (1400 rpm) for 2 hours at room temperature.

For immunoprecipitation of endogenous GBP1/GBP1pS156, 2 μg of antibodies against the proteins and 50 μl of protein G sepharose beads (ab193259, Abcam) were added per 1 ml of lysates (~2 mg of protein) and incubated on a rotator overnight at 4°C. Beads were washed with lysis buffer five times, and proteins were eluted by acidification with 50 μl of 0.2 M glycine pH 2.0 followed by an immediate wash with 50 μl of lysis buffer. Input, unbound, and elution fractions were kept during immunoprecipitations and analyzed by immunoblotting.

Semiquantitative immunoprecipitation

To determine the affinity of GBP1 binding to 14-3-3 proteins, we adapted a previously published

protocol for semiquantitative coimmunoprecipitation (58). For this, HA-tagged 14-3-3 proteins were produced by transfection of HEK293T cells and purified by HA immunoprecipitation. Proteins were eluted using HA peptide and concentrated using 10-kDa cutoff Amicon Ultra-15 Centrifugal Filters (UFC901024, Merck Millipore), exchanging the buffer with binding buffer (25 mM HEPES pH 7.3, 100 mM NaCl, 0.01% Triton X-100, 5% Glycerol, 1 mM DTT). Phosphorylated GBP1pS156 was purified from IFN-γ– and Dox-treated THP-1Δ*GBP1*+Tet-Flag-*GBP1* WT cells using the GBP1pS156 antibody, and the protein was concentrated with 50-kDa cutoff Amicon Ultra-15 Centrifugal Filters (UFC905024, Merck Millipore), exchanging the buffer with binding buffer. Next, the purified and phosphorylated Flag-GBP1 was bound to Flag(M2)-agarose beads (A2220, Sigma) by rotating at 4°C overnight and washed with binding buffer, and the beads were blocked with 50 mM ethanolamine in PBS. Finally, 10 μl of prepared and washed GBP1 bait beads were incubated with different concentrations of purified 14-3-3 proteins for 1 hour at 4°C and then washed five times with binding buffer. GBP1 and bound 14-3-3 proteins were eluted by boiling the beads in 2× Laemmli loading dye. Finally, obtained samples were analyzed by immunoblotting, and blots were quantified using FIJI/ImageJ.

MS analysis of GBP1 phosphorylation

In vitro kinase assay samples (with ATP and without ATP) were aliquoted in three technical replicates per condition and diluted with 8 M urea to 25 mM Tris-HCl, 4 M urea, 5 mM $MgCl_2$. Proteins were reduced with 5 mM Tris(2-carboxyethyl)phosphine-hydrochloride (TCEP) for 30 min at 37°C and alkylated with 10 mM iodoacetamide (IAA) for 30 min at room temperature in the dark. Subsequently, protein samples were loaded onto a 30-kDa molecular weight cutoff filter (MWCO) (Sartorius, VIVACON 500, VN01H22) and washed three times with 50 mM ammonium bicarbonate (Ambic) by centrifugation at 14,000*g*. Proteins were digested with GluC (Promega, V1651) in 50 mM Ambic at a protease/protein (w/w) ratio of 1:20 at 37°C overnight. Peptides were eluted by centrifugation at 14,000*g*, and the 30-kDa MWCO filter was washed once with 50 mM Ambic. Peptides were cleaned with C18 UltraMicroSpin columns (The Nest Group) following the manufacturer's recommendation and dried in a SpeedVac concentrator (Thermo Fisher Scientific).

Peptides were reconstituted in 3% acetonitrile (ACN)/0.1% formic acid (FA)/H_2O, and 0.5 μg per sample was subjected to liquid chromatography–MS/MS (LC-MS/MS) using an nLC1000 coupled to an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific). Peptides were loaded onto a 30-cm column (75-μm ID, New Objective) with mobile phase buffer A containing 0.1% FA/H_2O and separated by reverse-phase liquid chromatography at a ReproSil-Pur 120 A C18-AQ 1.- μm stationary phase (Dr. Maisch HPLC GmbH). Peptides were eluted with mobile phase buffer B, which consists of 99.9% ACN/0.1% FA, starting with 5% B and increasing up to 50% B in 100 min. LC-MS/MS measurements were acquired in data-dependent mode. MS1 spectra were recorded from 200 to 2000 *m/z* with a resolution of 120,000 at 200 *m/z* in the Orbitrap analyzer. MS1 scans were triggered at an AGC target of 1×10^6 and maximum injection time in auto mode. Within a cycle time of 3 s, peptides were iteratively isolated with a 1.3-*m/z* isolation window and fragmented by ETD with automatic charge-dependent determined ETD parameters. MS2 scans were recorded in the Orbitrap analyzer upon reaching the AGC target of 5×10^4 with injection time in auto mode. Fragment ions were monitored with a resolution of 60,000 at 200 *m/z*. Fragmented peptides were dynamically excluded from further analysis for 30 s. Ions with charge state unassigned, 1, or >6 were excluded from fragmentation.

Data-dependent acquisition (DDA) raw data were processed using MaxQuant (v1.5.6) (59). Spectra were searched against the sequences for recombinant GBP1 and PIM1, as well as the proteome sequences of *Escherichia coli* (strain B / BL21-DE3) (tax ID: 469008) and *Trichoplusia ni* (tax ID: 7111), all retrieved from UniProtKB (accessed on 30.05.2021), and common contaminations. Default parameters of MaxQuant were applied while choosing the protease GluC, allowing four miss cleavages, setting the variable modifications to Oxidation (M), Acetyl (Protein N-term), and Phospho (STY), and fixed modification to Carbamidomethyl (C). A false discovery rate (FDR) of 0.01 was controlled at peptide spectrum match, posttranslational modification (PTM) site, and protein level by MaxQuant. Phosphosite localization was automatically calculated by PTM Score of Andromeda within the MaxQuant proteomic suite. Phosphosite data were further analyzed with Perseus (v1.6.10.50). Phosphosites were filtered for a minimum localization probability of >0.75 and minimum three measurements in total across all conditions. A peaklist of ETD-activated MS/MS spectrum for GBP1pS156, preprocessed by Andromeda/MaxQuant, was exported from the MaxQuant viewer and annotated with Interactive Peptide Spectral Annotator (60), considering c-, z•-, and y-type fragment ions. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (61) partner repository with the dataset identifier PXD030010.

Identification of GBP1-interacting proteins by MS

Six-well plates were seeded with 10^7 THP-1Δ*GBP1*+Tet-Flag-*GBP1* cells , and the cells were differentiated, pretreated with IFN-γ and Dox, and infected with Tg. Cells were washed in ice-cold PBS, scraped from the plates, pelleted by centrifugation, and washed in PBS. Whole-cell lysates were prepared by adding 500 μl of lysis buffer (1% Triton X-100, 20 mM Tris–HCl pH 8.0, 130 mM NaCl, 1 mM DTT, 10 mM sodium fluoride) with added protease and phosphatase inhibitor cocktails and incubation for 15 min on ice. Lysates were cleared by high-speed centrifugation and used for Flag immunoprecipitation. The samples were run on a 12% Bis-Tris polyacrylamide gel until the running front had entered the gel ~5 mm. Samples were excised from the gel, destained with 50% acetonitrile/50 mM ammonium bicarbonate, reduced with 10 mM DTT, and alkylated with 55 mM IAA. After alkylation, proteins were digested with 250 ng of trypsin overnight at 37°C, and the peptides were extracted in 2% formic acid and 1% acetonitrile and speed vacuum dried.

The peptides were reconstituted in 50 μl of 0.1% trifluoroacetic acid (TFA) before analysis and loaded on a 50-cm EASY-Spray column (75-μm inner diameter, 2-μm particle size, Thermo Fisher Scientific) equipped with an integrated electrospray emitter. Reverse-phase LC was performed using the RSLC nano U3000 (Thermo Fisher Scientific) with a binary buffer system at a flow rate of 275 nl/min. Buffer A was 0.1% FA and 5% DMSO, and buffer B was 80% acetonitrile, 0.1% FA, and 5% DMSO. The samples were run on a linear gradient of buffer B (2 to 30%) in 95.5 min. The nano LC was coupled to an Orbitrap Fusion Lumos mass spectrometer using an EASY-Spray nano source (Thermo Fisher Scientific). The Orbitrap Fusion Lumos was operated in DDA mode acquiring MS1 scan (R = 120,000) in the Orbitrap, followed by higher-energy collisional dissociation (HCD) MS2 scans in the Ion Trap. Top Speed acquisition algorithm with a cycle time of 3 s was used to determine the number of selected precursor ions for fragmentation. The dynamic exclusion was set at 30 s. For ion accumulation, the MS1 target was set to 4×10^5 ions and the MS2 target to 2×10^3 ions. The maximum ion injection time used for MS1 scans was 50 ms and that for MS2 scans was 300 ms. The HCD normalized collision energy was set at 28, and the ability to inject ions for all available parallelizable time was set to "true."

Orbitrap .RAW files were analyzed by MaxQuant (version 1.6.0.13) using Andromeda for peptide search. For identification, peptide length was set to seven amino acids, and match between runs was enabled. Parent ion and tandem mass spectra were searched against UniprotKB *Homo sapiens* (tax ID: 9606) and Tg (tax ID: 1080348) databases. For the search, the enzyme specificity was set to trypsin, with a maximum of two missed cleavages. The precursor mass tolerance was set to 20 parts per million (ppm)

for the initial search (used for mass recalibration) and to 6 ppm for the main search. Product mass tolerance was set to 20 ppm. Fixed modification was defined as cysteine carbamidomethylation. Methionine oxidation and N-terminal protein acetylation were searched as variable modifications. The datasets were filtered on posterior error probability to achieve 1% FDR on the protein level. Quantification was performed with the LFQ algorithm in MaxQuant using three replicate measurements per experiment. Results were compared against the CRAPome database (*62*) to eliminate common contaminants of Flag immunoprecipitations (data S3 and S4). The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (*61*) partner repository with the dataset identifier PXD029463.

In vitro kinase assay

To determine in vitro kinase activity of PIM1 (PV3503, Thermo Fisher Scientific), PIM2 (PV3649, Thermo Fisher Scientific), PIM3 (A30516, Thermo Fisher Scientific), and Akt1/PKB (P2999, Thermo Fisher Scientific) against GBP1 (ab114960, Abcam), the recombinant proteins were mixed with 100 μM ATP (PV3227, Thermo Fisher Scientific) in kinase buffer (50 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$) and allowed to incubate for 30 min at 30°C. Reactions were stopped by addition of Laemmli buffer containing 5% DTT and denaturing at 95°C for 5 min before analysis by immunoblotting or silver stain.

GTPase assay

To determine GTPase activity, 2 μM GBP1, in vitro phosphorylated GBP1, or purified GBP1:14-3-3σ complex was incubated with 1 mM of GTP in reaction buffer (40 mM Tris pH 7.5, 80 mM NaCl, 8 mM $MgCl_2$, 1 mM EDTA) at room temperature. The concentration of free phosphate product was subsequently determined using Malachite Green Phosphate Assay Kit (MAK307, Sigma) by measuring absorbance at 620 nm and interpolating a phosphate standard curve.

Antibody binding assay

Enzyme-linked immunosorbent assay (ELISA) plates were coated with 1 μM phosphorylated or unphosphorylated GBP1 peptides in 100 μl of coating buffer (50 mM carbonate/bicarbonate pH 9.6) at 4°C overnight. Plates were washed with TBS-T and blocked with 1% BSA in TBS-T for 1 hour at room temperature. Plates were washed again, and GBP1pS156 antibody was added in different dilutions in blocking buffer and left to bind at 4°C overnight. Plates were washed and probed with horseradish peroxidase (HRP)–conjugated anti-rabbit secondary antibody and developed with TMB solution (N301, Thermo Fisher Scientific) for 10 min at room temperature. The reaction was quenched by the addition of 2 N H_2SO_4, and absorption was quantified on a plate reader (Fluostar Omega, BMG Labtech).

ITC

ITC was performed on a MicroCal VP-ITC. For each titration, 14-3-3σ and GBP1 or previously in vitro phosphorylated GBP1 were dialyzed into reaction buffer (50 mM HEPES pH 7.4, 150 mM KCl). Titrations were performed at 25°C with serial injections of 7-μl aliquots of 14-3-3σ (125 or 250 μM) into a solution of GBP1 or GBP1pS156 (12.5 or 25 μM, respectively) in the calorimeter cell (volume 1.3 ml). Data were analyzed with ORIGIN software (MicroCal).

Fixed immunofluorescence microscopy

For imaging, 0.25 × 10^6 cells were seeded on coverslips in 24-well plates. After differentiation, treatments, and infection, cells were washed three times with warm PBS before fixation and then fixed with 4% methanol-free formaldehyde (28906, Thermo Fisher Scientific) for 15 min at room temperature. Fixed cells were washed with PBS and kept at 4°C overnight. Fixed specimens were permeabilized with PermQuench buffer [0.2% (w/v) BSA and 0.02% (w/v) saponin in PBS] for 30 min at room temperature and then stained with primary antibodies for 1 hour at room temperature. After three washes with PBS, cells were incubated with secondary antibody and 1 μg/ml Hoechst 33342 (H3570, Invitrogen) diluted in PermQuench buffer for 1 hour at room temperature. Coverslips were extensively washed with PBS and once with ddH_2O and mounted using 5 μl of Mowiol. Specimens were imaged on a Leica SP5-inverted confocal microscope at 100× magnification.

High-throughput imaging

For infection analysis, 5 × 10^4 THP-1s were seeded per well of a black-wall, clear-bottom 96-well imaging plate (Thermo Fisher Scientific), differentiated, treated, and infected as described above. Following fixation with 4% methanol-free formaldehyde, specimens were permeabilized and stained with 1 μg/ml Hoechst 33342 and 2 μg/ml CellMask Deep Red plasma membrane stain (C10046, Invitrogen) at room temperature for 1 hour. After staining, the specimens were extensively washed with PBS and kept in PBS for imaging.

For recruitment analysis, the cells were prepared as described above but were seeded on black-wall, glass-bottom 96-well imaging plates CG 1.0 (130-098-264, MACS Miltenyi Biotec). After fixation, cells were permeabilized and stained with primary antibody diluted in PermQuench buffer for 1 hour at room temperature. After three washes with PBS, cells were incubated with the appropriated secondary antibody and 1 μg/ml Hoechst 33342 diluted in PermQuench buffer for 1 hour at room temperature. Then, the specimens were extensively washed with PBS and kept in PBS for imaging.

Imaging for infection analysis used a Cell Insight CX7 High-Content Screening Platform (Thermo Fisher Scientific) or a Celldiscoverer 7 (Zeiss) using 10× magnification. For recruitment analysis, plates were imaged on an Opera Phenix High-Content Screening System (PerkinElmer) or a Celldiscoverer 7 (Zeiss) using 20×/40× magnification. After image acquisition, the images were fed into the Host Response to Microbe Analysis (HRMAn) analysis pipeline (*63*, *64*).

Vacuole and parasite breakage assay

For quantification of Tg vacuole integrity, cells seeded in black-wall 96-well imaging plates were infected and treated as described above. One hour before fixation, 1 μg/ml HCS CellMask Deep Red (H32721, Invitrogen) was added to the culture medium to load the cytosol of host cells with this fluorescent dye. After fixation and staining with Hoechst 33342, plates were imaged at 20× magnification on a Celldiscoverer 7 (Zeiss). Fluorescence of the dye within detected Tg vacuoles was then analyzed using HRMAn. For analysis of parasite disruption, cells expressing GFP_{11} fragment were infected with type II (Pru) TgΔ*Hpt*+$GFP_{1\text{-}10}$–expressing parasites. After the infection time course, cells were fixed, stained with Hoechst and anti-Tg-SAG1, imaged, and analyzed for GFP fluorescence using HRMAn, as described previously (*13*).

Fluorescence recovery after photobleaching

Quantification of GBP1 recruitment dynamics was performed using fluorescence recovery after photobleaching (FRAP). For this, 0.1 × 10^6 THP-1 cells were seeded per well of an eight-well ibidi μ-slide (80826, ibidi) differentiated, treated, and infected with Tg as described above. At 2 hours after infection, the specimens were moved to a Leica SP5-inverted confocal microscope. The experiment was performed using the LAS software FRAP-wizard with the following parameters: 256 by 256 pixels, 1400-Hz line frequency scan speed with bidirectional scan, 2 Airy unit (AU) pinhole size, and bleaching laser power at 100%. After selection of the bleach area on Tg vacuoles, the experiment was performed with the following time course: 10 frames at 120 ms prebleach, 1 frame at 120 ms bleach, 100 frames at 120 ms postbleach I, 10 frames at 1 s postbleach II, and 10 frames at 5 s postbleach III. After acquisition, data was double normalized (*65*) to correct for acquisition bleaching. Finally, using curve fit, the half-time of recovery was determined as a measurement for the mobility of the molecules.

Imaging quantifications

Most quantifications of images were performed using HRMAn (*63*, *64*). HRMAn was used for quantification of Tg growth, Golgi fragmentation, recruitment of proteins to pathogen vacuoles, and measurement of cell shapes, sizes and fluorescence properties. The Pearson's

correlation coefficient for colocalization analysis was computed using Coloc2 FIJI plugin (66). The average distance of GBP1 aggregates to the nucleus was determined using AggreCount (67). Golgi localization of GBP1 was quantified by determining the mean fluorescence intensity (MFI) of GBP1 in the $GM130^+$ area and divided by the MFI of the respective cell cytosol to normalize for uneven expression.

Recombinant protein expression and purification

GBP1 and *14-3-3*σ ORFs were amplified from pLenti-Tet-*GBP1* and pLenti-Tet-*14-3-3*σ and ligated into NdeI- and BamHI-digested pET15d vector, pretagged N-terminally with an MBP-8×His-3C or 6×His-3C sequence using Gibson assembly to obtain pET15d-MBP-8×His-3C-*GBP1* and pET15d-6×His-3C-*14-3-3*σ, respectively. Plasmids were transformed into BL21 *E. coli*. Bacteria from a single colony were picked and grown in a 10 ml of LB plus ampicillin (EU0400-D, Euromedex) starter culture overnight. Starter cultures were used to inoculate 12 liters of LB plus ampicillin cultures (1:1000 dilution), which were grown to OD_{600} ~ 0.9. Protein expression was induced by the addition of 1 μM β-D-1-thiogalactopyranoside (IPTG) (EU0008-B, Euromedex). GBP1 was expressed for 16 hours at 18°C and 14-3-3σ at 30°C for 4 hours while shaking at 150 rpm.

Bacteria were harvested by centrifugation, washed with PBS, resuspended in 100 ml of lysis buffer (GBP1: 50 mM HEPES pH 7.8, 30 mM imidazole, 150 mM KCl; 14-3-3σ: 50 mM NaH_2PO_4 pH 8.0, 10 mM imidazole; 300 mM NaCl), and lysed by sonication on ice (12 min, 1-s pulses with 1-s interruptions). The lysate was cleared by high-speed centrifugation at 20,000*g* for 1 hour at 4°C and cleared lysate loaded onto a 5-ml Ni-NTA column (ab270529, Abcam). The column was washed with 20 ml of lysis buffer, and bound proteins were eluted with a 40-ml gradient from 30 to 500 mM imidazole in lysis buffer. Target protein–containing fractions were pooled and dialyzed overnight against size exclusion buffer [size exclusion chromatography (SEC) buffer: 50 mM HEPES pH 7.8, 150 mM KCl] while also digesting with His-3C protease (Protein Expression and Purification Core Facility, EMBL Heidelberg). His-3C protease and cleaved tags were removed using reverse immobilized metal affinity chromatography Ni-IMAC purification (635660, Takara), and proteins were concentrated using 50-kDa cutoff spin column concentrators (Amicon Ultra).

The full-length PIM1 ORF was amplified from pLenti-Tet-*PIM1* and ligated into BamHI- and HindIII-digested pLIB vector modified with a N-terminal MBP-TEV-SBP tag using Gibson assembly to obtain pLIB-MBP-TEV-SBP-*PIM1*. The assembled vector was used for Baculovirus generation in SF21 cells (SF21: 11497013, Gibco; SF-900 II SFM: 11497013, Gibco) as described previously (68, 69). Five milliliters of Baculovirus was used for transfection of 500 ml of Hi5 (*T. ni* High Five Cells: B85502, Invitrogen; ExpressFive SFM, 10486025, Gibco) culture and incubated at 27°C for 72 hours. Hi5 insect cells were harvested by centrifugation, washed with PBS, resuspended in lysis buffer (50 mM HEPES pH 8.0, 300 mM NaCl), and lysed by sonication at 4°C (12 min, 1-s pulses with 1-s interruptions). The lysate was cleared by high-speed centrifugation at 20,000*g* for 1 hour at 4°C. Cleared lysate was loaded onto preequilibrated amylose resin (E8021S, New England Biolabs) and left to bind for 4 hours at 4°C. The resin was washed with 10 ml of lysis buffer three times, and bound PIM1 was eluted with 30 ml of lysis buffer containing 25 mM maltose. Obtained MBP-TEV-SBP-PIM1 was dialyzed against 50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, and 30 mM imidazole buffer overnight while also digesting with His-TEV protease (Protein Expression and Purification Core Facility, EMBL Heidelberg). His-TEV protease and cleaved tags were removed using serial reverse His- and amylose-affinity purification and PIM1 protein concentrated using 30-kDa cutoff spin column concentrators (Amicon Ultra). Bulk-purified GBP1, PIM1, and 14-3-3σ were further purified by SEC with a HiLoad 16/600 Superdex 200 column (Cytiva) in SEC buffer.

GBP1:14-3-3σ complex formation and purification

For in vitro reconstitution of the GBP1:14-3-3σ complex, purified GBP1, 14-3-3σ, and PIM1 were mixed in a 2:4:1 ratio in complex formation buffer (Tris-HCl pH 7.4, 150 mM KCl, 10 mM $MgCl_2$) with addition of 1 mM ATP and incubated at 30°C for 1 hour. Reaction samples were cleared by centrifugation at 20,000*g* for 5 min at 4°C and ran over a Superdex 200 Increase 10/300 GL column (Cytiva Lifesciences) in SEC buffer. To further polish for structure analysis, complex-containing fractions were pooled, concentrated with a 100-kDa cutoff spin column concentrator (Amicon Ultra), and repurified by SEC. To test the effect of 14-3-3σ binding on GBP1 dimerization, samples were additionally incubated with 1 mM GTP or with 1 mM GDP plus 10 mM AlF_3 (449628, Merck).

Single-particle cryo-EM and structure modeling

The GBP1:14-3-3σ complex was prepared as described above. Fractions corresponding to a heterotrimeric complex were vitrified in R1.2/R1.3 UltrAuFoil grids with a 300 mesh (Q350AR13A, Electron Microscopy Sciences) that were glow discharged for 20 s at 25 mA and 0.3 bar using a Pelco EasyGlow device. The sample concentration was adjusted to 0.25 mg/ml, and 2-μl drops in SEC buffer were applied to either side of the grid. Excess sample was removed by blotting using a Vitribot MARK IV at 4°C and 100% humidity for 2 s with blotting force 10 and subsequently plunge-frozen in liquid ethane. The vitrified particles were imaged using a FEI Titan Krios cryo–transmission electron microscope (TEM) at 300 kV at EMBL Heidelberg equipped with a GIF Quantum K2 direct electron detector and a GIF quantum energy filter (Gatan). In total, 14,974 movies were collected in counting mode with seven exposures per hole at a nominal magnification of 215,000× with a pixel size of 0.638 $Å^2$. Each movie was exposed for 8 s, resulting in a total dose of 65.5 $e/Å^2$ over 40 frames.

All data-processing steps were performed using cryoSPARC v3.2.0 unless stated otherwise (70). Imported movies were subjected to motion correction using the cryoSPARC multipatch motion correction followed by multipatch contrast transfer function (CTF) estimation. Initially, particles were identified by using automated Gaussian picking for elliptical particles with 50 by 120 Å. Picked particles were extracted, fourfold binned, and subjected to extensive two-dimensional (2D) classification, which identified 4129 particles corresponding to the 14-3-3σ/GBP1 complex. The particles were used to train Topaz (71) and repick on all micrographs, yielding 420,768 particles. 2D classification identified 29,150 particles. Ab initio modeling with three models was used to generate preliminary volumes. Reextraction followed by a heterogeneous refinement using the previously determined initial 3D volumes resulted in a similar particle distribution. Class 3 was used to retrain Topaz for automated particle picking. Identified particles were extracted with fourfold binning. Subsequent 2D classification identified 138,809 particles, which were used to generate five new initial models. The particles were reextracted with full pixel size and subjected to heterogeneous refinement with the previously determined 3D volumes. Class 3 yielded a model with 40,112 particles. The corresponding particles were further classified into three ab initio models. A class with 18,717 particles was identified and subjected to nonuniform refinement, yielding a final cryo-EM map at 5.1 Å.

Densities were generated in CryoSPARC and loaded into ChimeraX (version 0.93) (72), and previously reported crystal structures [Protein Data Bank (PDB) IDs 1F5N for GBP1 and 1YWT for 14-3-3σ dimer) were fitted using the program's own density fitting tools. The final model was deposited in the Electron Microscopy Data Bank (EMDB) with accession number EMD-18149, and the rigid body docked model was deposited in the PDB with accession number 8Q4L.

Evolutionary analysis of GBP genes

Position-Specific Iterated-Basic Local Alignment Search Tool (PSI-BLAST) (73) searches using full-length GBP1 as query of the RefSeq Protein sequence database identified >3300 GBP1-like proteins from >500 species. Protein sequences were aligned with MAFFT (74), and incomplete

data were removed with MaxAlign v1.1 (*75*). The closest homolog of human GBP1 could be identified based on sequence homology in 484 species, and the respective sequences were analyzed for the presence of the corresponding phosphoserine and PIM1 recognition motif.

In silico analysis, data handling, and statistics

Data were plotted using Prism 8.4.0 (GraphPad, Inc.) and presented as means of multiple experiments with error bars as standard deviation (SD), unless stated otherwise. The significance of results was determined by nonparametric one-way analysis of variance (ANOVA), unpaired Student's *t* test, or two-way ANOVA as indicated in the figure legends. Benjamini, Krieger, and Yekutieli FDR (Q = 5%)–based correction for multiple comparisons as implemented in Prism was used when making more than three comparisons. Structures of GBP1 were obtained from the PDB database under PDB IDs 1F5N and 1DG3 (*76*, *77*), and images were rendered with MacPymol v.1.74, which was also used for B-factor analysis. Molecular graphics and analyses were performed with UCSF ChimeraX. GBP1 phosphorylation-site data were obtained from www.phosphosite.org (accessed 4 January 2019). 14-3-3 binding site identification used 14-3-3-Pred (*78*). Kinase predictions were performed using NetPhos Server3.1 (*79*) and GPS5.0 (*80*), and recognition sequence phosphorylation-likelihood values were obtained from PhosphoNET: Kinexus (http://www.phosphonet.ca/). All open-source KNIME workflows used in this work can be found at https://github.com/HRMAn-Org/HRMAn and on the homepage https://www.hrman.org.

REFERENCES AND NOTES

1. J. D. MacMicking, Interferon-inducible effector mechanisms in cell-autonomous immunity. *Nat. Rev. Immunol.* **12**, 367–382 (2012). doi: 10.1038/nri3210; pmid: 22531325
2. E. A. Van der Biezen, J. D. G. Jones, Plant disease-resistance proteins and the gene-for-gene concept. *Trends Biochem. Sci.* **23**, 454–456 (1998). doi: 10.1016/S0968-0004(98)01311-5; pmid: 9868361
3. N. Lopes Fischer, N. Naseer, S. Shin, I. E. Brodsky, Effector-triggered immunity and pathogen sensing in metazoans. *Nat. Microbiol.* **5**, 14–26 (2020). doi: 10.1038/s41564-019-0623-2; pmid: 31857733
4. K. Tretina, E. S. Park, A. Maminska, J. D. MacMicking, Interferon-induced guanylate-binding proteins: Guardians of host defense in health and disease. *J. Exp. Med.* **216**, 482–500 (2019). doi: 10.1084/jem.20182031; pmid: 30755454
5. B.-H. Kim *et al*., A family of IFN-γ–inducible 65-kD GTPases protects against bacterial infection. *Science* **332**, 717–721 (2011). doi: 10.1126/science.1201711; pmid: 21551061
6. A. R. Shenoy *et al*., GBP5 promotes NLRP3 inflammasome assembly and immunity in mammals. *Science* **336**, 481–485 (2012). doi: 10.1126/science.1217141; pmid: 22461501
7. M. P. Wandel *et al*., Guanylate-binding proteins convert cytosolic bacteria into caspase-4 signaling platforms. *Nat. Immunol.* **21**, 880–891 (2020). doi: 10.1038/s41590-020-0697-2; pmid: 32541830
8. M. A. Al-Zeer, H. M. Al-Younes, D. Lauster, M. Abu Lubad, T. F. Meyer, Autophagy restricts *Chlamydia trachomatis* growth in human macrophages via IFNG-inducible guanylate binding proteins. *Autophagy* **9**, 50–62 (2013). doi: 10.4161/auto.22482; pmid: 23086406
9. A. S. Piro *et al*., Detection of cytosolic *Shigella flexneri* via a C-terminal triple-arginine motif of GBP1 inhibits actin-based motility. *mBio* **8**, e01979-17 (2017). doi: 10.1128/mBio.01979-17; pmid: 29233899
10. A. Qin *et al*., Guanylate-binding protein 1 (GBP1) contributes to the immunity of human mesenchymal stromal cells against Toxoplasma gondii. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 1365–1370 (2017). doi: 10.1073/pnas.1619665114; pmid: 28123064
11. D. Fisch *et al*., Human GBP1 is a microbe-specific gatekeeper of macrophage apoptosis and pyroptosis. *EMBO J.* **38**, e100926 (2019). doi: 10.15252/embj.2018100926; pmid: 31268602
12. A. R. Bass, S. Shin, Human GBP1 promotes pathogen vacuole rupture and inflammasome activation during *Legionella pneumophila* infection. bioRxiv 2020.05.27.120477 [Preprint] (2020); https://doi.org/10.1101/2020.05.27.120477.
13. D. Fisch *et al*., Human GBP1 differentially targets *Salmonella* and *Toxoplasma* to license recognition of microbial ligands and caspase-mediated death. *Cell Rep.* **32**, 108008 (2020). doi: 10.1016/j.celrep.2020.108008; pmid: 32783936
14. M. Kutsch *et al*., Direct binding of polymeric GBP1 to LPS disrupts bacterial cell envelope functions. *EMBO J.* **39**, e104926 (2020). doi: 10.15252/embj.2020104926; pmid: 32510692
15. J. C. Santos *et al*., Human GBP1 binds LPS to initiate assembly of a caspase-4 activating platform on cytosolic bacteria. *Nat. Commun.* **11**, 3276 (2020). doi: 10.1038/s41467-020-16889-z; pmid: 32581219
16. N. Britzen-Laurent *et al*., Intracellular trafficking of guanylate-binding proteins is regulated by heterodimerization in a hierarchical manner. *PLOS ONE* **5**, e14246 (2010). doi: 10.1371/journal.pone.0014246; pmid: 21151871
17. E. Kravets *et al*., Guanylate binding proteins directly attack *Toxoplasma gondii* via supramolecular complexes. *eLife* **5**, e11479 (2016). doi: 10.7554/eLife.11479; pmid: 26814575
18. S. Zhu *et al*., Cryo-ET of a human GBP coatomer governing cell-autonomous innate immunity to infection. bioRxiv 2021.08.26.457804 [Preprint] (2021); https://doi.org/10.1101/2021.08.26.457804.
19. M. Schwemmle, P. Staeheli, The interferon-induced 67-kDa guanylate-binding protein (hGBP1) is a GTPase that converts GTP to GMP. *J. Biol. Chem.* **269**, 11299–11305 (1994). doi: 10.1016/S0021-9258(19)78125-3; pmid: 7512561
20. N. Modiano, Y. E. Lu, P. Cresswell, Golgi targeting of human guanylate-binding protein-1 requires nucleotide binding, isoprenylation, and an IFN-γ-inducible cofactor. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 8680–8685 (2005). doi: 10.1073/pnas.0503227102; pmid: 15937107
21. M. Andreoli *et al*., Identification of the first inhibitor of the GBP1:PIM1 interaction. Implications for the development of a new class of anticancer agents against paclitaxel resistant cancer cells. *J. Med. Chem.* **57**, 7916–7932 (2014). doi: 10.1021/jm5009902; pmid: 25211704
22. A. Kumar *et al*., Crystal structures of proto-oncogene kinase Pim1: A target of aberrant somatic hypermutations in diffuse large cell lymphoma. *J. Mol. Biol.* **348**, 183–193 (2005). doi: 10.1016/j.jmb.2005.02.039; pmid: 15808862
23. M. T. Yip-Schneider, M. Horie, H. E. Broxmeyer, Transcriptional induction of pim-1 protein kinase gene expression by interferon gamma and posttranscriptional effects on costimulation with steel factor. *Blood* **85**, 3494–3502 (1995). doi: 10.1182/blood.V85.12.3494.bloodjournal85123494; pmid: 7540064
24. T. L. T. Aho *et al*., Pim-1 kinase promotes inactivation of the pro-apoptotic Bad protein by phosphorylating it on the Ser^{112} gatekeeper site. *FEBS Lett.* **571**, 43–49 (2004). doi: 10.1016/j.febslet.2004.06.050; pmid: 15280015
25. A. N. Bullock, J. Debreczeni, A. L. Amos, S. Knapp, B. E. Turk, Structure and substrate specificity of the Pim-1 kinase. *J. Biol. Chem.* **280**, 41675–41682 (2005). doi: 10.1074/jbc.M510711200; pmid: 16227208
26. A. Macdonald *et al*., Pim kinases phosphorylate multiple sites on Bad and promote 14-3-3 binding and dissociation from Bcl-XL. *BMC Cell Biol.* **7**, 1 (2006). doi: 10.1186/1471-2121-7-1; pmid: 16403219
27. C. K. Palaty, I. Clark-Lewis, D. Leung, S. L. Pelech, Phosphorylation site substrate specificity determinants for the Pim-1 protooncogene-encoded protein kinase. *Biochem. Cell Biol.* **75**, 153–162 (1997). doi: 10.1139/o97-026; pmid: 9250363
28. N. Ostler *et al*., Gamma interferon-induced guanylate binding protein 1 is a novel actin cytoskeleton remodeling factor. *Mol. Cell. Biol.* **34**, 196–209 (2014). doi: 10.1128/MCB.00664-13; pmid: 24190970
29. P. Riou *et al*., 14-3-3 proteins interact with a hybrid prenyl-phosphorylation motif to inhibit G proteins. *Cell* **153**, 640–653 (2013). doi: 10.1016/j.cell.2013.03.044; pmid: 23622247
30. G. Gay *et al*., *Toxoplasma gondii* TgIST co-opts host chromatin repressors dampening STAT1-dependent gene regulation and IFN-γ-mediated host defenses. *J. Exp. Med.* **213**, 1779–1798 (2016). doi: 10.1084/jem.20160340; pmid: 27503074
31. P. Olias, R. D. Etheridge, Y. Zhang, M. J. Holtzman, L. D. Sibley, *Toxoplasma* effector recruits the Mi-2/NuRD complex to repress STAT1 transcription and block IFN-γ-dependent gene expression. *Cell Host Microbe* **20**, 72–82 (2016). doi: 10.1016/j.chom.2016.06.006; pmid: 27414498
32. G. Carbotti *et al*., Cytokine-induced guanylate binding protein 1 (GBP1) release from human ovarian cancer cells. *Cancers* **12**, 488 (2020). doi: 10.3390/cancers12020488; pmid: 32093058
33. J. Ma, H. K. Arnold, M. B. Lilly, R. C. Sears, A. S. Kraft, Negative regulation of Pim-1 protein kinase levels by the B56β subunit of PP2A. *Oncogene* **26**, 5145–5153 (2007). doi: 10.1038/sj.onc.1210323; pmid: 17297438
34. L.-M. Ting, A. C. Kim, A. Cattamanchi, J. D. Ernst, *Mycobacterium tuberculosis* inhibits IFN-γ transcriptional responses without inhibiting activation of STAT1. *J. Immunol.* **163**, 3898–3906 (1999). doi: 10.4049/jimmunol.163.7.3898; pmid: 10490990
35. P. Najarro, P. Traktman, J. A. Lewis, Vaccinia virus blocks gamma interferon signal transduction: Viral VH1 phosphatase reverses Stat1 activation. *J. Virol.* **75**, 3185–3196 (2001). doi: 10.1128/JVI.75.7.3185-3196.2001; pmid: 11238845
36. D. M. Miller *et al*., Human cytomegalovirus inhibits major histocompatibility complex class II expression by disruption of the Jak/Stat pathway. *J. Exp. Med.* **187**, 675–683 (1998). doi: 10.1084/jem.187.5.675; pmid: 9480977
37. D. Nandan, N. E. Reiner, Attenuation of gamma interferon-induced tyrosine phosphorylation in mononuclear phagocytes infected with *Leishmania donovani*: Selective inhibition of signaling through Janus kinases and Stat1. *Infect. Immun.* **63**, 4495–4500 (1995). doi: 10.1128/iai.63.11.4495-4500.1995; pmid: 7591091
38. K. Kumthip *et al*., Hepatitis C virus NS5A disrupts STAT1 phosphorylation and suppresses type I interferon signaling. *J. Virol.* **86**, 8581–8591 (2012). doi: 10.1128/JVI.00533-12; pmid: 22674974
39. J. J. He *et al*., Global transcriptome profiling of multiple porcine organs reveals toxoplasma gondii-induced transcriptional landscapes. *Front. Immunol.* **10**, 1531 (2019). doi: 10.3389/fimmu.2019.01531; pmid: 31333663
40. S.-Z. Wu *et al*., Genome-wide CRISPR screen identifies host factors required by *Toxoplasma gondii* infection. *Front. Cell. Infect. Microbiol.* **9**, 460 (2020). doi: 10.3389/fcimb.2019.00460; pmid: 32039045
41. L. Zhang, J. Chen, H. Fu, Suppression of apoptosis signal-regulating kinase 1-induced cell death by 14-3-3 proteins. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8511–8515 (1999). doi: 10.1073/pnas.96.15.8511; pmid: 10411906
42. R. Zhang *et al*., Hsp90-Akt phosphorylates ASK1 and inhibits ASK1-mediated apoptosis. *Oncogene* **24**, 3954–3963 (2005). doi: 10.1038/sj.onc.1208548; pmid: 15782121
43. J. J. Gu, Z. Wang, R. Reeves, N. S. Magnuson, PIM1 phosphorylates and negatively regulates ASK1-mediated apoptosis. *Oncogene* **28**, 4261–4271 (2009). doi: 10.1038/onc.2009.276; pmid: 19749799
44. D. Morishita, R. Katayama, K. Sekimizu, T. Tsuruo, N. Fujita, Pim kinases promote cell cycle progression by phosphorylating and down-regulating p27Kip1 at the transcriptional and posttranscriptional levels. *Cancer Res.* **68**, 5076–5085 (2008). doi: 10.1158/0008-5472.CAN-08-0634; pmid: 18593906
45. R. A. L. van der Hoorn, S. Kamoun, From guard to decoy: A new model for perception of plant pathogen effectors. *Plant Cell* **20**, 2009–2017 (2008). doi: 10.1105/tpc.108.060194; pmid: 18723576
46. I. Jéru *et al*., Interaction of pyrin with 14.3.3 in an isoform-specific and phosphorylation-dependent manner regulates its translocation to the nucleus. *Arthritis Rheum.* **52**, 1848–1857 (2005). doi: 10.1002/art.21050; pmid: 15934090
47. W. Gao, J. Yang, W. Liu, Y. Wang, F. Shao, Site-specific phosphorylation and microtubule dynamics control Pyrin inflammasome activation. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E4857–E4866 (2016). doi: 10.1073/pnas.1601700113; pmid: 27482109
48. S. L. Masters *et al*., Familial autoinflammation with neutrophilic dermatosis reveals a regulatory mechanism of pyrin activation. *Sci. Transl. Med.* **8**, 332ra45 (2016). doi: 10.1126/scitranslmed.aaf1471; pmid: 27030597
49. Y. H. Park, G. Wood, D. L. Kastner, J. J. Chae, Pyrin inflammasome activation and RhoA signaling in the autoinflammatory diseases FMF and HIDS. *Nat. Immunol.* **17**, 914–921 (2016). doi: 10.1038/ni.3457; pmid: 27270401
50. M. M. Gaidt *et al*., Self-guarding of MORC3 enables virulence factor-triggered immunity. *Nature* **600**, 138–142 (2021). doi: 10.1038/s41586-021-04054-5; pmid: 34759314

51. M. H. Orzalli *et al*., Virus-mediated inactivation of anti-apoptotic Bcl-2 family members promotes Gasdermin-E-dependent pyroptosis in barrier epithelial cells. *Immunity* **54**, 1447–1462.e5 (2021). doi: 10.1016/j.immuni.2021.04.012; pmid: 33979579
52. K. Wanigasooriya *et al*., Patient derived organoids confirm that PI3K/AKT signalling is an escape pathway for radioresistance and a target for therapy in rectal cancer. *Front. Oncol.* **12**, 920444 (2022). doi: 10.3389/fonc.2022.920444; pmid: 35860583
53. N. E. Sanjana, O. Shalem, F. Zhang, Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014). doi: 10.1038/nmeth.3047; pmid: 25075903
54. Z. Dwight, R. Palais, C. T. Wittwer, uMELT: Prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. *Bioinformatics* **27**, 1019–1020 (2011). doi: 10.1093/bioinformatics/btr065; pmid: 21300699
55. E. W. Wilker, R. A. Grant, S. C. Artim, M. B. Yaffe, A structural basis for 14-3-3σ functional specificity. *J. Biol. Chem.* **280**, 18891–18898 (2005). doi: 10.1074/jbc.M500982200; pmid: 15731107
56. F. Q. Li, A. Mofunanya, K. Harris, K. Takemaru, Chibby cooperates with 14-3-3 to regulate β-catenin subcellular distribution and signaling activity. *J. Cell Biol.* **181**, 1141–1154 (2008). doi: 10.1083/jcb.200709091; pmid: 18573912
57. J. Lugrin, F. Martinon, Detection of ASC oligomerization by Western blotting. *Bio Protoc.* **7**, ••• (2017). doi: 10.21769/BioProtoc.2292; pmid: 29085861
58. S. Lapetina, H. Gil-Henn, A guide to simple, direct, and quantitative in vitro binding assays. *J. Biol. Methods* **4**, e62 (2017). doi: 10.14440/jbm.2017.161; pmid: 31453222
59. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008). doi: 10.1038/nbt.1511; pmid: 19029910
60. D. R. Brademan, N. M. Riley, N. W. Kwiecien, J. J. Coon, Interactive Peptide Spectral Annotator: A versatile web-based tool for proteomic applications. *Mol. Cell. Proteomics* **18**, S193–S201 (2019). doi: 10.1074/mcp.TIR118.001209; pmid: 31088857
61. Y. Perez-Riverol *et al*., The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019). doi: 10.1093/nar/gky1106; pmid: 30395289
62. D. Mellacheruvu *et al*., The CRAPome: A contaminant repository for affinity purification-mass spectrometry data. *Nat. Methods* **10**, 730–736 (2013). doi: 10.1038/nmeth.2557; pmid: 23921808
63. D. Fisch *et al*., Defining host-pathogen interactions employing an artificial intelligence workflow. *eLife* **8**, e40560 (2019). doi: 10.7554/eLife.40560; pmid: 30744806
64. D. Fisch *et al*., HRMAn 2.0: Next-generation artificial intelligence-driven analysis for broad host-pathogen interactions. *Cell. Microbiol.* **23**, e13349 (2021). doi: 10.1111/cmi.13349; pmid: 33930228
65. M. Kang, M. Andreani, A. K. Kenworthy, Validation of normalizations, scaling, and photofading corrections for FRAP data analysis. *PLOS ONE* **10**, e0127966 (2015). doi: 10.1371/journal.pone.0127966; pmid: 26017223
66. J. Schindelin *et al*., Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012). doi: 10.1038/nmeth.2019; pmid: 22743772
67. J. A. Klickstein, S. Mukkavalli, M. Raman, AggreCount: An unbiased image analysis tool for identifying and quantifying cellular aggregates in a spatially defined manner. *J. Biol. Chem.* **295**, 17672–17683 (2020). doi: 10.1074/jbc.RA120.015398; pmid: 33454006
68. I. Berger, D. J. Fitzgerald, T. J. Richmond, Baculovirus expression system for heterologous multiprotein complexes. *Nat. Biotechnol.* **22**, 1583–1587 (2004). doi: 10.1038/nbt1036; pmid: 15568020
69. C. Bieniossek, T. Imasaki, Y. Takagi, I. Berger, MultiBac: Expanding the research toolbox for multiprotein complexes. *Trends Biochem. Sci.* **37**, 49–57 (2012). doi: 10.1016/j.tibs.2011.10.005; pmid: 22154230
70. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017). doi: 10.1038/nmeth.4169; pmid: 28165473
71. T. Bepler *et al*., Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019). doi: 10.1038/s41592-019-0575-8; pmid: 31591578
72. E. F. Pettersen *et al*., UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021). doi: 10.1002/pro.3943; pmid: 32881101
73. S. F. Altschul *et al*., Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997). doi: 10.1093/nar/25.17.3389; pmid: 9254694
74. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013). doi: 10.1093/molbev/mst010; pmid: 23329690
75. R. Gouveia-Oliveira, P. W. Sackett, A. G. Pedersen, MaxAlign: Maximizing usable data in an alignment. *BMC Bioinformatics* **8**, 312 (2007). pmid: 17725821
76. B. Prakash, L. Renault, G. J. Praefcke, C. Herrmann, A. Wittinghofer, Triphosphate structure of guanylate-binding protein 1 and implications for nucleotide binding and GTPase mechanism. *EMBO J.* **19**, 4555–4564 (2000). doi: 10.1093/emboj/19.17.4555; pmid: 10970849
77. B. Prakash, G. J. K. Praefcke, L. Renault, A. Wittinghofer, C. Herrmann, Structure of human guanylate-binding protein 1 representing a unique class of GTP-binding proteins. *Nature* **403**, 567–571 (2000). doi: 10.1038/35000617; pmid: 10676968
78. F. Madeira *et al*., 14-3-3-Pred: Improved methods to predict 14-3-3-binding phosphopeptides. *Bioinformatics* **31**, 2276–2283 (2015). doi: 10.1093/bioinformatics/btv133; pmid: 25735772
79. N. Blom, T. Sicheritz-Pontén, R. Gupta, S. Gammeltoft, S. Brunak, Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4**, 1633–1649 (2004). doi: 10.1002/pmic.200300771; pmid: 15174133
80. C. Wang *et al*., GPS 5.0: An update on the prediction of kinase-specific phosphorylation sites in proteins. *Genomics Proteomics Bioinformatics* **18**, 72–80 (2020). doi: 10.1016/j.gpb.2020.01.001; pmid: 32200042

ACKNOWLEDGMENTS

We thank all members of the Frickel lab (Crick and University of Birmingham) and the Galej lab (EMBL Grenoble) for helpful discussions. We dedicate this work to the memory of Leo Mercer. We further thank H. Ploegh, J. Langhorne, J. Kagan, C. Reis e Sousa, and C. Bunce for critical comments on the manuscript. We thank M. Goldberg for providing essential reagents. We thank G. Clark, V. Dearing (Crick Genomics and Equipment Park), M. Howell, O.-R. Song (Crick High Throughput Screening), and the Crick and EMBL Grenoble lab support staff for their help. We thank M. Pelosse for assistance with imaging and managing the EMBL Grenoble eukaryotic expression facility (EEF), S. Schneider for the EM facility support at EMBL Grenoble, F. Weis (EMBL cryo-EM service platform in Heidelberg) for performing data collection, and C. Mas and the biophysical platforms of the Grenoble Instruct-ERIC center (UAR 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology. We also thank the Jinek group (University of Zürich) for providing IT infrastructure for cryo-EM data processing. **Funding:** This research was funded, in whole or in part, by The Wellcome Trust. D.F. was supported by a Boehringer Ingelheim Fonds PhD fellowship, a Human Frontier Science Program (HFSP) long-term postdoctoral fellowship (LT0006/2022-L), and an EMBO nonstipendiary postdoctoral fellowship (ALTF 491-2022). M.Y. acknowledges support from the Agency for Medical Research and Development (AMED) Strategic International Collaborative Research Program (JP20jm0210067), Research Program on Emerging and Re-emerging Infectious Diseases (JP20fk0108137), and Japanese Initiative for Progress of Research on Infectious Diseases for Global Epidemic (JP20wm0325010); and a Grant-in-Aid for Transformative Research Area (B) (20B304), for Scientific Research on Innovation Areas (19H04809), for Scientific Research (B) (18KK0226 and 18H02642), and for Scientific Research (A) (19H00970) from the Ministry of Education, Culture, Sports, Science and Technology, Fusion Oriented Research for Disruptive Science and Technology. J.M. was supported by the European Research Council (649101-UbiProPox) and by Core funding to MRC Laboratory for Molecular Cell Biology at University College London (MC_UU_00012/7). A.R.S. was supported by MRC grant MR/P022138/1. E.-M.F. received funding from Wellcome Trust Senior Research Fellowship 217202/Z/19/Z, MRC Research and Innovation Grant MR/V030930/1 (also to A.R.S.), MRC-AMED grant MR/T029323/1 (also to M.Y.), and core funding to the Francis Crick Institute by Cancer Research UK, the UK Medical Research Council, and the Wellcome Trust (FC001076). K.M.M. acknowledges support from CRUK CEA grant C61638/A27112. A.D.B. is funded by a CRUK Advanced Clinician Scientist Award (C31641/A23923). The Grenoble Instruct-ERIC center (UAR 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology is supported by FRISBI (ANR-10-INBS-0005-02) and GRAL, which are financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). A.P.S. received core funding to the Francis Crick Institute by Cancer Research UK, the UK Medical Research Council, and the Wellcome Trust (FC001999). **Author contributions:** D.F. and E.-M.F. conceived the study. D.F. performed most of the biochemical and functional experiments. D.F., M.M.P., E.A., and X.L. performed protein expression and purification. M.M.P. and X.L. performed single-particle cryo-EM and data analysis and built the structure. F.W. performed MS for mapping of GBP1 phosphorylation sites. V.E. performed MS analysis to identify GBP1 interactomes. K.M.M., G.M.M., and A.D.B. provided essential reagents and performed organoid culture experiments. H.B. and M.Y. provided essential reagents. E.A., B.C., and S.L.-R. provided essential experimental protocols, assistance, and valuable suggestions. A.P.S., A.R.S., J.M., B.W., W.P.G., and E.-M.F. acquired and contributed funding and experimental oversight. D.F. and E.-M.F. analyzed the data. D.F., E.A., A.R.S., and E.-M.F. wrote the manuscript with input from all authors. All authors discussed the results and commented on the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. For underlying data, see data S1 to S6. The proteomics raw data have been deposited to PRIDE (PXD029463 and PXD030010). The cryo-EM data have been deposited on EMDB (EMD-18149). All specific and stable reagents generated in this study are available from the corresponding author without restriction. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This research was funded in whole or in part by The Francis Crick Institute (FC001076 and FC001999), which is funded by Wellcome a cOAlition S organization. The author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg2253
Figs. S1 to S16
Tables S1 to S6
Reference (*81*)
MDAR Reproducibility Checklist
Data S1 to S6

Submitted 14 December 2022; resubmitted 3 July 2023
Accepted 23 August 2023
10.1126/science.adg2253

RESEARCH ARTICLE SUMMARY

CORONAVIRUS

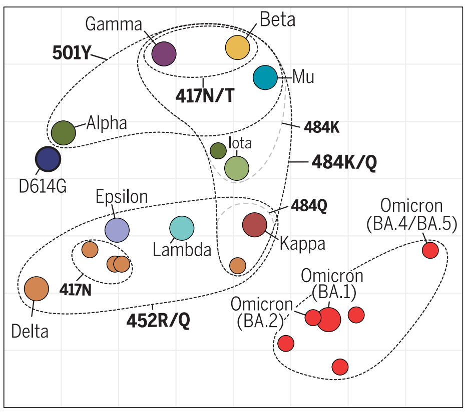
Mapping SARS-CoV-2 antigenic relationships and serological responses

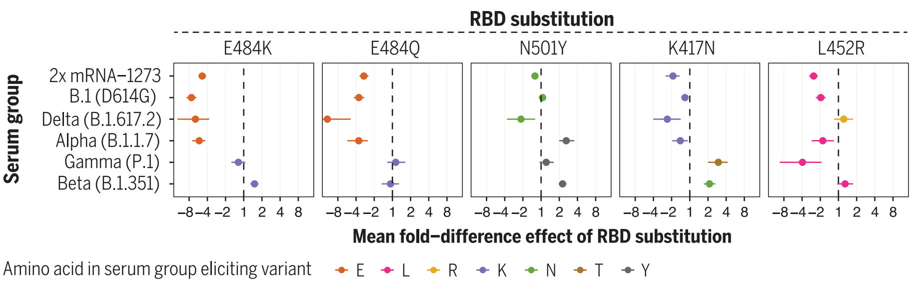
Samuel H. Wilks*†, Barbara Mühlemann†, Xiaoying Shen†, Sina Türeli, Eric B. LeGresley, Antonia Netzl, Miguela A. Caniza, Jesus N. Chacaltana-Huarcaya, Victor M. Corman, Xiaoju Daniell, Michael B. Datto, Fatimah S. Dawood, Thomas N. Denny, Christian Drosten, Ron A. M. Fouchier, Patricia J. Garcia, Peter J. Halfmann, Agatha Jassem, Lara M. Jeworowski, Terry C. Jones, Yoshihiro Kawaoka, Florian Krammer, Charlene McDanal, Rolando Pajon, Viviana Simon, Melissa S. Stockwell, Haili Tang, Harm van Bakel, Vic Veguilla, Richard Webby, David C. Montefiori*, Derek J. Smith*

INTRODUCTION: Vaccination has greatly reduced the disease burden of SARS-CoV-2. However, since late 2020, variants have emerged that are able to escape immunity from vaccination and previous infections, including B.1.351 (Beta), P.1 (Gamma), B.1.617.2 (Delta), and B.1.1.529 (Omicron and its descendants). In combination with vaccination, infections with different variants form the basis of current population immunity against SARS-CoV-2.

RATIONALE: Understanding the antigenic relationships among SARS-CoV-2 variants, the substitutions that cause antigenic change, and how the immune response is shaped by previous infections, is crucial for understanding the evolution of the virus, determining whether new variants avoid neutralization from antibodies induced by current vaccines, and evaluating the need for vaccine updates.

To this end, we analyzed patterns of neutralization and cross-reactivity among a panel of 21 SARS-CoV-2 variants and 15 groups of human sera obtained from individuals after primary infection with one of 10 different variants or after D614G or B.1.351 vaccination. First, we sought to understand patterns of cross-reactivity and response breadth in postvaccination responses 4 weeks and >3 months after second or third vaccine doses. Then, we used antigenic cartography to visualize antigenic relationships between 21 SARS-CoV-2 variants and experimentally test point mutations to investigate the drivers of the antigenic changes observed in the antigenic map. Lastly, we investigated how serological reactivity postinfection relates to the primary-exposure variant.





Serological responses following exposure to different SARS-CoV-2 variants. (Top) Antigenic map showing antigenic relatedness of SARS-CoV-2 variants. Variants are grouped by amino acid substitutions present at positions 417, 452, 484, and 501, with an additional grouping for the six Omicron variants. (Bottom) Average effect size of single substitutions at different RBD positions on sera from individuals with different primary exposures, colored by the amino acid present at the respective RBD position in the eliciting variant.

RESULTS: Quantifying changes in cross-reactivity and response breadth after vaccination, our results show the largest increase between 4 weeks and >3 months after a second dose. In particular, we found that the main short-term effect of the third vaccination was to boost the magnitude of a response that had already become more cross-reactive rather than to generate significant additional breadth of cross-reactivity.

Using antigenic cartography, we inferred and subsequently experimentally tested our inference that antigenic differences among pre-Omicron variants are primarily caused by substitutions at spike-protein positions 417, 452, 484, and 501 (see figure, top). The experimental effect of these substitutions was largely consistent with those inferred from the map.

We also found that sensitivity to these substitutions varied greatly between individuals infected with different variants. These differences are consistent with substantial changes in immunodominance of different spike regions, depending on the variant an individual was first exposed to and the amino acid present at these positions in the eliciting variant (see figure, bottom). For example, whereas sera of individuals exposed to D614G, B.1.1.7, and B.1.617.2 variants are sensitive to changes at position 484, sera of individuals exposed to B.1.351 and P.1 are not.

CONCLUSION: Our results provide a comprehensive analysis of the antigenic variation between SARS-CoV-2 variants and the development of the immune response after infection or vaccination. The large antigenic effect of a small number of substitutions in the receptor-binding domain (RBD) of SARS-CoV-2 is similar to the pattern observed for seasonal influenza viruses, for which major antigenic changes are often associated with single or double substitutions. These substitutions in the SARS-CoV-2 RBD not only allow the virus to escape from preexisting immunity but also influence the regions of the spike protein that the immune response targets. Depending on their infection history, different individuals can thus be sensitive to substitutions in different regions of the spike protein. As individuals increasingly experience multiple infections, choosing vaccine immunogens on the basis of immunodominance considerations may be an important aspect in ensuring high vaccine efficacy across populations with different patterns of preexisting immunity. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: sw463@cam.ac.uk (S.H.W.); david.montefiori@duke.edu (D.C.M.); dsmith@zoo.cam.ac.uk (D.J.S.)
†These authors contributed equally to this work.
Cite this article as S. H. Wilks *et al.*, *Science* **382**, eadj0070 (2023). DOI: 10.1126/science.adj0070

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adj0070

RESEARCH ARTICLE

CORONAVIRUS

Mapping SARS-CoV-2 antigenic relationships and serological responses

Samuel H. Wilks[1]*†, Barbara Mühlemann[2,3]†, Xiaoying Shen[4,5]†, Sina Türeli[1], Eric B. LeGresley[1], Antonia Netzl[1], Miguela A. Caniza[6], Jesus N. Chacaltana-Huarcaya[7], Victor M. Corman[2,3], Xiaoju Daniell[4], Michael B. Datto[8], Fatimah S. Dawood[9], Thomas N. Denny[5], Christian Drosten[2,3], Ron A. M. Fouchier[10], Patricia J. Garcia[11], Peter J. Halfmann[12], Agatha Jassem[13], Lara M. Jeworowski[2], Terry C. Jones[1,2,3], Yoshihiro Kawaoka[12,14,15,16], Florian Krammer[17,18], Charlene McDanal[4], Rolando Pajon[19], Viviana Simon[17,18,20,21], Melissa S. Stockwell[22], Haili Tang[4], Harm van Bakel[23], Vic Veguilla[9], Richard Webby[24], David C. Montefiori[4,5]*, Derek J. Smith[1]*

During the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic, multiple variants escaping preexisting immunity emerged, causing reinfections of previously exposed individuals. Here, we used antigenic cartography to analyze patterns of cross-reactivity among 21 variants and 15 groups of human sera obtained after primary infection with 10 different variants or after messenger RNA (mRNA)–1273 or mRNA-1273.351 vaccination. We found antigenic differences among pre-Omicron variants caused by substitutions at spike-protein positions 417, 452, 484, and 501. Quantifying changes in response breadth over time and with additional vaccine doses, our results show the largest increase between 4 weeks and >3 months after a second dose. We found changes in immunodominance of different spike regions, depending on the variant an individual was first exposed to, with implications for variant risk assessment and vaccine-strain selection.

Since the beginning of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic, the virus has caused more than 766 million cases and 6.9 million deaths (*1*). During the first year of the pandemic, circulation was dominated by the B.1 variant, characterized by the D614G substitution in the spike protein, which imparted increased infectivity and transmissibility in vitro and in animal models (*2*, *3*) but did not escape serum neutralization (*4*). Since then, multiple variants have circulated widely, with five of them—B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), B.1.617.2 (Delta), and B.1.1.529 (Omicron and descendant sublineages)—categorized as variants of concern by the World Health Organization (WHO) on the basis of evidence of higher transmissibility, increased virulence, and/or reduced effectiveness of vaccines, therapeutics, or diagnostics (*5*). Sublineages arising within the Omicron lineage show considerable antigenic diversity, often being as antigenically different from the early Omicron variant (Omicron BA.1) as BA.1 is from the B.1-like viruses (*6*).

Before the emergence of the Omicron variants, several variants circulating widely in 2021 were antigenically distinct from the B.1-like 2020 viruses, with B.1.351 escaping neutralization by convalescent and postvaccination sera the most (*7*, *8*). Moreover, sera from individuals first infected with B.1.351 and P.1 failed to readily neutralize B.1.617.2, and vice versa, although B.1.617.2 did not show strong escape from convalescent sera after infection with prototype-like variants (*9*, *10*). All variants in the Omicron lineage show substantial escape from postvaccination and convalescent sera after infection with the pre-Omicron variants (*11*, *12*).

Understanding the antigenic relationships among SARS-CoV-2 variants and the substitutions that cause antigenic change is required for evaluating the need for vaccine updates and predicting whether new variants avoid neutralization from antibodies induced by current vaccines. However, with an increasing number of variants, understanding antigenic relationships through neutralization titer data becomes more complex. Antigenic cartography (*13*) is a tool originally developed for the analysis of human seasonal influenza virus antigenic data and has since been used in the analysis of antigenic variation in other pathogens, including avian, equine, and swine influenza viruses (*14*–*18*); flaviviruses (*19*), including dengue viruses (*20*, *21*); lyssaviruses (*22*); and foot-and-mouth disease viruses (*23*). It provides a quantitative and visual summary of antigenic differences among large numbers of variants and is a core component of the biannual influenza virus vaccine-strain selection process convened by the WHO. In this study, we used antigenic cartography to analyze patterns of cross-reactivity among a panel of 21 SARS-CoV-2 variants and 15 groups of human sera obtained from individuals after primary infection with one of 10 different variants or after prototype or B.1.351 primary-series vaccination. We followed this with experimental testing of point mutations to investigate the drivers of the antigenic changes observed and how the effects of subsequent changes on serological reactivity relate to the primary-exposure variant.

Results

Variant reactivity by serum group

We used 207 serum specimens collected from vaccinated or infected individuals (table S1) and titrated in a Food and Drug Administration (FDA)–approved (FDA/CBER Master File 026862) neutralization assay against lentiviral pseudotypes encoding the spike protein of 21 SARS-CoV-2 variants (fig. S1). Titrations included a panel of 15 pre-Omicron variants and 6 Omicron variants. Before the collection of each serum sample, individuals had reported no known previous SARS-CoV-2 infections or vaccination (table S1). The infecting variant was determined by using whole-genome sequencing. The serum specimens are from individuals with infections with D614G (n = 15), B.1.1.7 (Alpha, n = 14), B.1.351 (Beta, n = 19),

[1]Center for Pathogen Evolution, Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK. [2]Institute of Virology, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, 10117 Berlin, Germany. [3]German Centre for Infection Research (DZIF), partner site Charité, 10117 Berlin, Germany. [4]Department of Surgery, Duke University School of Medicine, Durham, NC, USA. [5]Duke Human Vaccine Institute, Duke University School of Medicine, Durham, NC, USA. [6]Department of Global Pediatric Medicine, Department of Infectious Diseases, St. Jude Children's Research Hospital, Memphis, TN, USA. [7]Hospital Nacional Daniel A. Carrión, Callao, Bellavista, Peru. [8]Department of Pathology, Duke University School of Medicine, Durham, NC, USA. [9]Centers for Disease Control and Prevention, Atlanta, GA, USA. [10]Erasmus Medical Center, Rotterdam, Netherlands. [11]School of Public Health, Universidad Peruana Cayetano Heredia, Lima, Peru. [12]Influenza Research Institute, Department of Pathobiological Science, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, WI, USA. [13]BC Centre for Disease Control, Vancouver, British Columbia, Canada. [14]Division of Virology, Institute of Medical Science, University of Tokyo, Tokyo, Japan. [15]The Research Center for Global Viral Diseases, National Center for Global Health and Medicine Research Institute, Tokyo, Japan. [16]Pandemic Preparedness, Infection and Advanced Research Center (UTOPIA), University of Tokyo, Tokyo, Japan. [17]Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [18]Department of Pathology, Cellular and Molecular Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [19]Moderna, Inc., Cambridge, MA, USA. [20]Division of Infectious Diseases, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [21]The Global Health and Emerging Pathogen Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [22]Division of Child and Adolescent Health, Department of Pediatrics, Columbia University Vagelos College of Physicians and Surgeons, and Department of Population and Family Health, Mailman School of Public Health, New York, NY, USA. [23]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [24]Department of Infectious Diseases, St. Jude Children's Research Hospital, Memphis, TN, USA.
***Corresponding author. Email: sw463@cam.ac.uk (S.H.W.); david.montefiori@duke.edu (D.C.M.); dsmith@zoo.cam.ac.uk (D.J.S.)**
†These authors contributed equally to this work.

P.1 (Gamma, *n* = 17), B.1.617.2 (Delta, *n* = 28), B.1.526+E484K (Iota, *n* = 6), B.1.637 (*n* = 3), C.37 (Lambda, *n* = 4), BA.1 (Omicron, *n* = 7), and BA.2 (Omicron, *n* = 1). We also included groups of sera collected from individuals twice vaccinated with a vaccine comprising prototypic SARS-CoV-2 spike (Moderna mRNA-1273) at 4 weeks (*n* = 32) or >3 months (*n* = 16) after a second dose, and sera after a third dose of the same vaccine at 4 weeks (*n* = 26) or >3 months (*n* = 8). Lastly, we included samples collected from individuals 4 weeks after a second dose of a B.1.351 spike vaccine (Moderna mRNA-1273.351) (*n* = 11).

The geometric mean titer (GMT) and individual reactivity profiles for 183 sera after the exclusion of 24 outliers with titers indicative of an unreported previous infection are shown in Fig. 1 (see "Excluding outlier sera" section in Materials and methods, and fig. S2). Each serum group exhibited a distinct profile of neutralization against the tested variants and, as expected, homologous serum and virus pairs were among the most potent in each group



Fig. 1. Neutralization of lentivirus pseudotypes encoding different SARS-CoV-2 spike proteins against different groups of human sera collected after vaccination or primary infection with different variants. Serum groups are split into (**A**) sera elicited by infection with different variants and (**B**) sera elicited by vaccination. Variants are ordered according to geometric mean titer (GMT) in D614G sera [(A), top left], and additional Omicron variants are ordered chronologically. Bold lines with empty circles show the GMTs calculated after estimated differences in individual response size were removed to mitigate biases in which not all sera from a group were titrated against a particular variant (as described in Materials and methods "Titer analyses" section). Fainter individual lines and solid points show individual serum titers. Points in the gray region at the bottom of the plots show titers and GMTs that fell below the detection threshold of 20. Each panel is labeled according to the respective serum group and color coded as indicated on the *x* axis. Figure S3 shows titers split by sample source for the five serum groups in which samples came from a mixture of cohorts or agencies. Titer box plots, line plots showing the individual serum titers after accounting for individual effects, and titer fold differences relative to the homologous titer are shown in figs. S4, S5, and S6, respectively. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. In the mutants, other amino acids were substituted at certain locations; for example, D614G indicates that aspartic acid at position 614 was replaced with glycine.

(although some serum groups included only a small number of sera) (Fig. 1A). The B.1.351 and P.1 serum groups both exhibited similar cross-neutralization of the B.1.351 and P.1 variants, suggesting a shared phenotype consistent with the closely matched receptor-binding domain (RBD) substitutions in these two variants (fig. S1). The Omicron variants showed the greatest escape from sera after second and third mRNA-1273 vaccination. Notably, BA.4/BA.5 titers were less than the other Omicron variants tested in the 4-weeks post–third mRNA-1273 vaccination sera but not in the samples taken directly before a third dose (>3 months after second mRNA-1273) or >3 months after a third dose. Comparable escape was seen for the Omicron and B.1.617.2 variants against sera obtained after two doses of mRNA-1273.351. Most BA.1 sera, and the BA.2 serum sample, showed no detectable reactivity to any pre-Omicron variants. However, for one BA.1 serum sample, in which titers were very high, some low levels of measurable reactivity were measurable for the B.1.351 and D614G variants (Fig. 1A). Overall, our findings relating to relative antigenic escape of the different Omicron variants against different serum groups were consistent with findings in other studies that used both lentivirus pseudotype and live-virus neutralization assays, in which overlap was present (*12*) (fig. S7).

Comparing postvaccination and postinfection sera, less variation among individuals was seen in the mRNA-1273 and mRNA-1273.351 vaccine sera than in the corresponding D614G and B.1.351 convalescent sera, both in terms of response magnitude and the pattern of reactivity seen against the variants (Fig. 1 and fig. S5). This is likely related to the more standardized dose (*24*) and nature of vaccination compared with infection (*25*–*27*) but might also be due to varying time intervals between infection and specimen collection in the convalescent specimens compared with those in postvaccination specimens (*28*).

Comparing vaccine response breadth

We calculated the breadth of postinfection and postvaccination responses, controlling for differences in titer magnitude by focusing on changes in the pattern of fold drops for each variant relative to the homologous variant (see "Calculating fold-drop differences in vaccine sera" in Materials and methods). We found that titer fold drops relative to D614G in each of the mRNA-1273 vaccination serum groups had a pattern similar to those of D614G convalescent sera, but the size of the fold drops was decreased, corresponding to an increased response breadth (Fig. 2A). Moreover, we found a temporal pattern of increasing response breadth (Fig. 2B), with the largest increase between 4 weeks and >3 months after a second dose and, to a lesser extent, between 4 weeks and >3 months after a third dose. In samples taken 4 weeks after a third vaccine dose, although titers were strongly boosted, breadth remained very similar to that measured in >3 months after two-dose mRNA-1273 samples taken directly before the third dose. For the B.1.351 postinfection and postvaccination groups, although titers were higher in the mRNA-1273.351 vaccination group compared with those of B.1.351 convalescent sera, we did not find evidence for a significant difference in the breadth of cross-reactivity (fig. S8).

Antigenic cartography

To visualize and quantify how the different variants relate to each other antigenically, we used the titrations shown in Fig. 1 to construct an antigenic map, in which antigens and sera are positioned relative to each other such that the distance between them corresponds to the fold drop compared with the maximum serum titer (see "Antigenic cartography" in Materials and methods). To incorporate the information from the different post–mRNA-1273 vaccine serum groups but also account for how their increased cross-reactivity would otherwise underestimate the relative antigenic differences between variants, we scaled distance estimates from these serum groups according to the fold-change difference estimates shown in Fig. 2B. Cross-validation results indicated that the neutralization data could generally be well represented in two dimensions (fig. S9), as shown in Fig. 3A. Overall, the antigenic relationships depicted in this map were robust to assay noise and the exclusion of serum groups and variants (Materials and methods, and figs. S10 to S21). The antigenic distinction between the B.1.617.2 variant and the three B.1.617.2 variants with K417N was, however, found to be predominantly driven by patterns of reactivity in the B.1.617.2 sera (figs. S16 and S18).

The clearest deviation from a two-dimensional (2D) representation of the antigenic relationships was due to the BA.4/BA.5 variant titers, for which antigenic differences fit better in three dimensions (Fig. 3B). Compared with its position in 2D space, the variant occupies a position that is more antigenically distinct from the other Omicron variants and closer to the B.1.617.2 sera. This positioning of BA.4/BA.5 reflects the fold change between BA.1 and BA.4/BA.5 in the B.1.617.2 and BA.1 serum groups (fig. S25). BA.4/BA.5 showed a drop in titers compared with BA.1 in the BA.1 convalescent sera but showed increased titers compared with BA.1 in the B.1.617.2 convalescent sera, possibly because BA.4/BA.5 and B.1.617.2 share the substitutions L452R and T478K. The relative antigenic distances estimated for BA.4/BA.5 were quite robust in relation to pre-Omicron serum groups, but the lower number of Omicron sera did lead to some additional uncertainty in the antigenic distance relative to the BA.1 and BA.2 sera.

Serological reactivity shown by antibody landscapes

Antigenic relationships depicted in the antigenic map provide a summary of how the reactivity of the different serum groups distributes among the variants. Figure 4 extends this visualization to antibody landscapes, where a surface in a third dimension represents an estimate of how the reactivity of each serum group and individual serum varies across antigenic space (see "Construction of the antibody landscapes" in Materials and methods). The x-y plane is depicted by the antigenic map in Fig. 3A, and the height of the landscape over a particular point or variant represents the estimated magnitude of serum reactivity in that antigenic region. Antibody landscapes therefore give an indication of how serum reactivity distributes after exposure to different variants, show how the magnitudes of the responses compare, and predict expected levels of reactivity to variants that have not been titrated (fig. S26).

Consistent with the titers in Fig. 1, the shape of the antibody landscapes is similar for the D614G and B.1.1.7 serum groups, with the highest reactivity centered on the D614G and B.1.1.7 variants. Similarly, the shape of the landscapes generated by the mRNA-1273.351, B.1.351, and P.1 sera is comparable, with highest reactivity centered on the B.1.351 and P.1 variants. The landscape of the B.1.617.2 sera shows a contrasting topology, with the highest reactivity indicated against the B.1.617.2 variant and falling off toward other areas of the map. On average, the B.1.617.2 sera titers against non-B.1.617.2 variants were often lower than predicted from the landscape. This suggests either that the B.1.617.2 sera could discriminate between B.1.617.2 and other variants more than was seen in reverse in the non-B.1.617.2 serum groups or that our measurements of B.1.617.2 serum antibodies against B.1.617.2 were biased toward higher values (fig. S26). Because our data showed a larger difference in reactivity between the B.1.617.2 and D614G variants in B.1.617.2 sera when compared with other sources (fig. S7), we speculate that the latter possibility may be the case. Separately, and in agreement with the titers, the landscapes show that the pre-Omicron sera investigated here would be expected to have markedly reduced reactivity against variants in the Omicron lineage (Fig. 4 and fig. S26). However, we note again that titers against the BA.4/BA.5 variant are less reliably predicted when assuming antigenic relationships in only two dimensions. Landscapes of BA.1 first-infection sera show that the drop-off of titers to largely nondetectable levels against pre-Omicron variants is steeper than that seen in reverse for the pre-Omicron sera landscapes. This could be related to the small number of

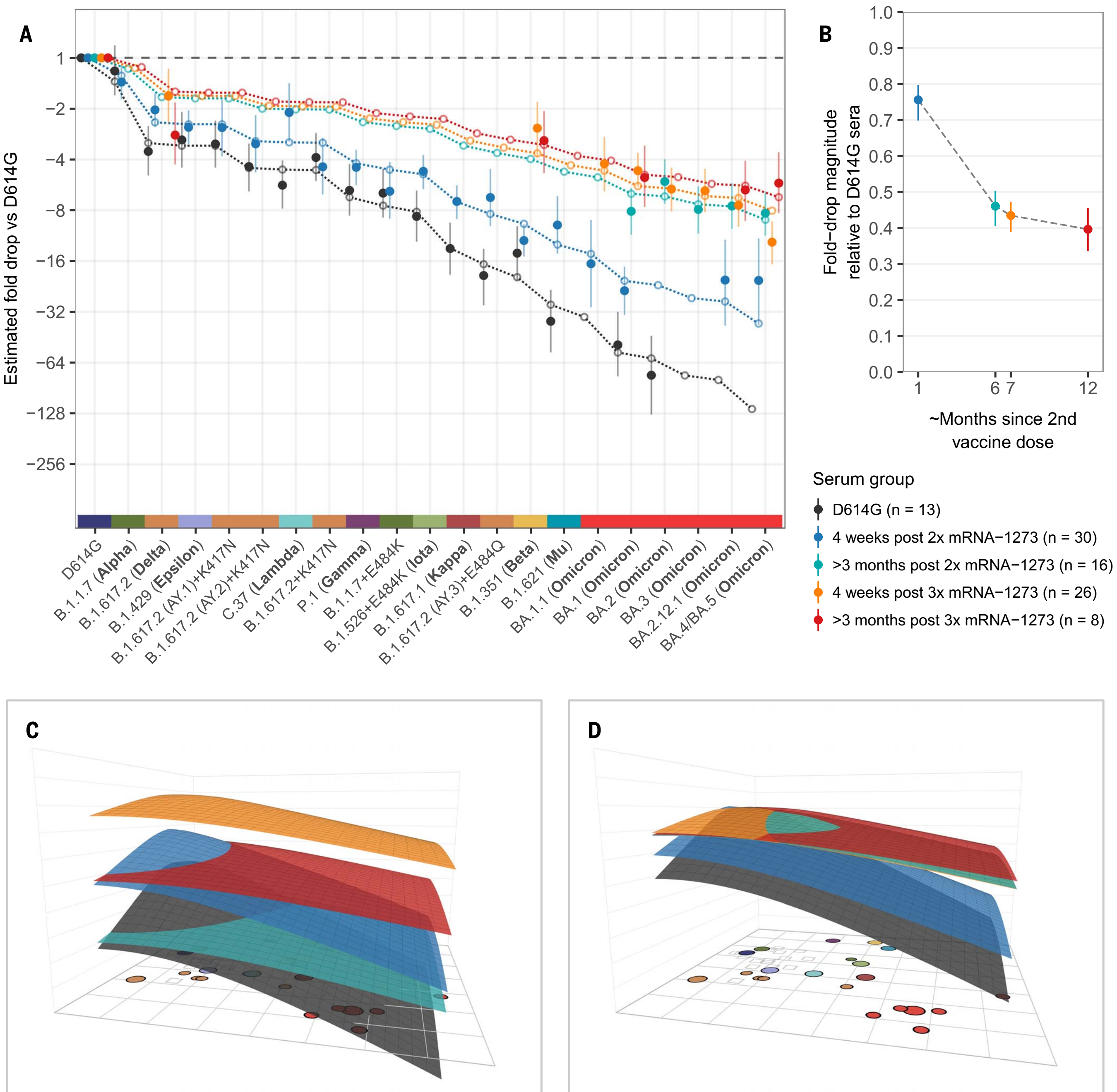


Fig. 2. Comparison of fold drops with different variants in post–D614G infection and post–mRNA-1273 vaccination sera. (**A**) Comparison of different estimates of titer fold-drop responses against different variants. Solid points show the estimate for the mean fold drop compared with the titer for D614G, and solid lines represent the 95% HDI for this estimate. The points for D614G to the left of the plot represent the homologous virus against which fold change for other strains was compared and are therefore fixed at 1. Dotted lines and outline circles show estimates based on a model that assumes a shared overall pattern of fold drops but estimates "slope" differences in the rate of reactivity drop-off seen in the four serum groups (as described in Materials and methods, "Calculating fold-drop differences in vaccine sera"). To aid comparison, points and lines for each of the serum groups are slightly offset from each other in the x axis to ensure that points do not directly overlap each other. (**B**) Estimates of fold-drop magnitudes for each mRNA-1273 serum group, relative to the fold-drops seen in the D614G convalescent serum group. Lines show the 95% HDI for each of the estimates, and the position on the *x* axis is proportional to the number of months since a second vaccine dose, assuming an average of 6 months for sera in the >3-months post–2× mRNA-1273 and >3-months post–3× mRNA-1273 groups. (**C**) Antibody landscapes showing how estimates of the mean titer for each of the serum groups in (A) vary across antigenic space. (**D**) Antibody landscapes, as shown in (C), but with a fixed peak titer of 2560 against the D614G variant to highlight differences in titer drop-off slopes. Interactive versions of the landscapes shown in (C) and (D) are available at https://acorg.github.io/mapping_SARS-CoV-2_antigenic_relationships_and_serological_responses.

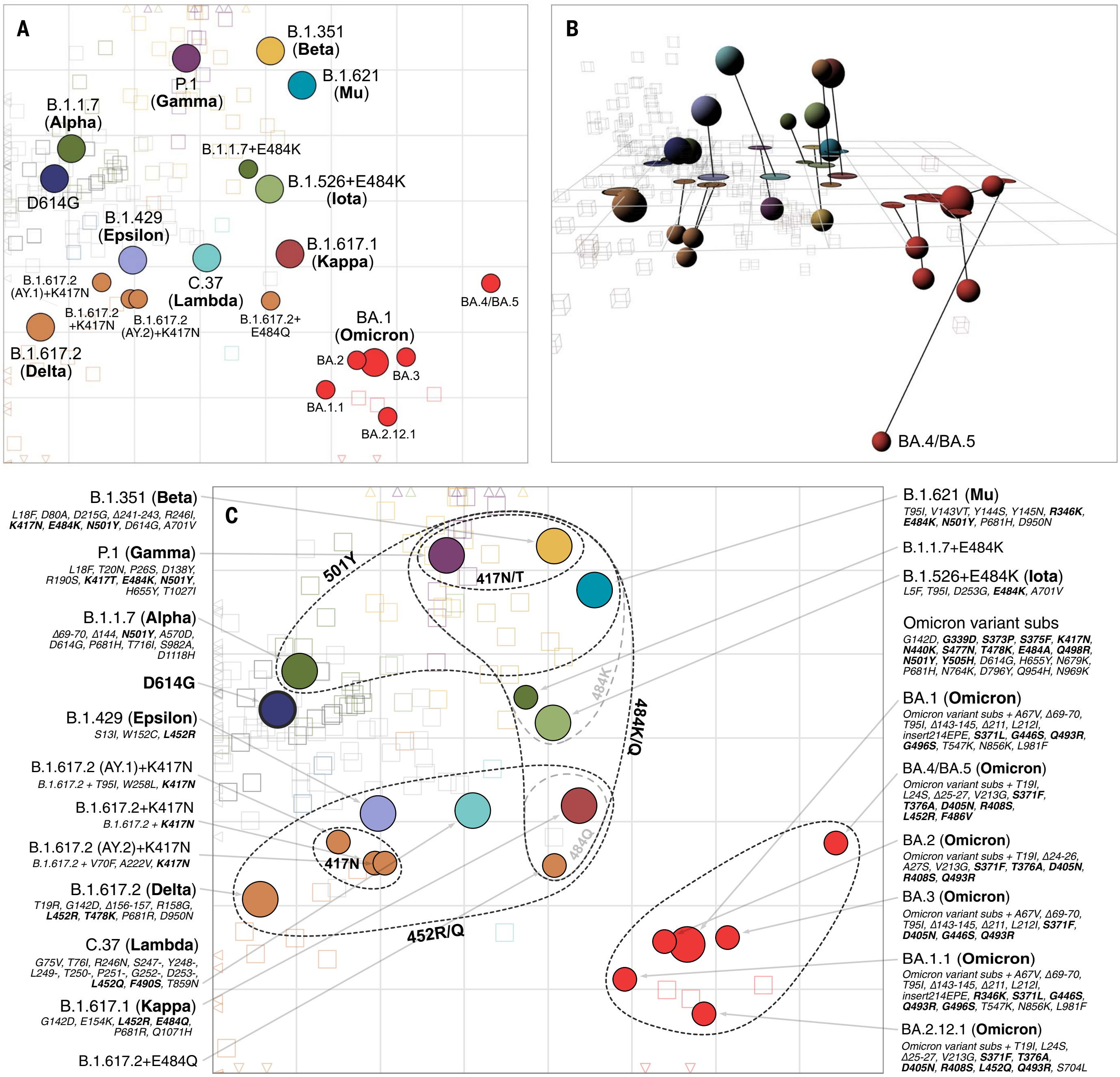


Fig. 3. Antigenic map of SARS-CoV-2 variants and selected substitutions. (**A**) Antigenic map with variant names. (**B**) Antigenic map with variant positions in 3D representation and lines connecting to their respective positions in the 2D map. (**C**) Antigenic map with variant names and substitutions annotated and grouped by amino acid present at spike positions 417, 452, 484, and 501, with an additional grouping for the six Omicron variants. Variants are shown as circles, sera as squares. Variants with additional substitutions from a root variant are denoted by smaller circles in the color of their root variant. The *x* and *y* axes both represent antigenic distance, with one grid square corresponding to a twofold serum dilution in the neutralization assay. Therefore, two grid squares correspond to a fourfold dilution, three to an eightfold dilution, and so on. The *x*-*y* orientation of the map is free because only the relative distances between variants and sera are relevant. Triangular arrowheads at the edge of the bounding box point in the direction of the sera that would be shown outside of the plot limits. A nonzoomed version of this map is shown in fig. S24. Interactive versions of the maps shown in (A) and (B) are available at https://acorg.github.io/mapping_SARS-CoV-2_antigenic_relationships_and_serological_responses.

BA.1 sera or to inherent asymmetries in BA.1 and pre-Omicron sera cross-reactivity. In general, the result is consistent with other results showing that BA.1 infections generate low titers to pre-Omicron variants (*29*, *30*).

The landscapes for different mRNA-1273 postvaccination sera confirm expectations for how cross-reactivity differs depending on the number of, and time since, vaccinations. These differences can largely be modeled by different slopes of titer reduction across antigenic space, with reactivity that peaks in the same antigenic region but decreases at differing rates (Fig. 4, K to M, and Fig. 2, B and C). The breadth of increase is clear at >3 months after a second

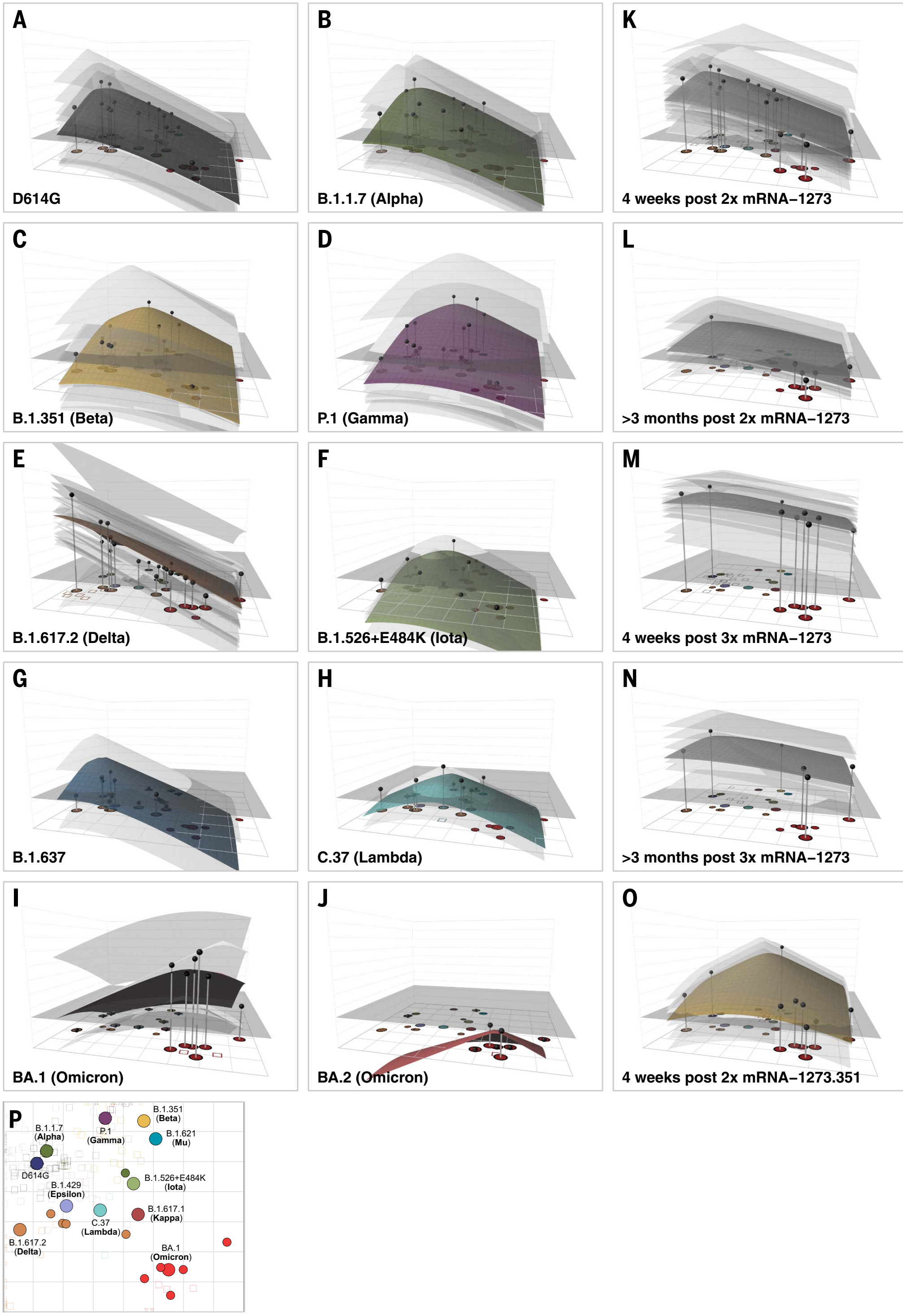


Fig. 4. Antibody landscapes for each serum group. Colored surfaces show the GMT antibody landscapes for the different serum groups; light-gray surfaces show the landscapes for individual sera. Gray impulses (vertical lines) show the height of the GMT for a specific variant, after accounting for individual effects as described in Materials and methods (which would otherwise bias the GMT for variants not titrated against all sera). The base *x-y* plane and plane colors correspond to the antigenic map shown in Fig. 3 and reproduced in (**P**). The vertical *z* axis in each plot corresponds to the titer on the $\log_2$ scale; each twofold increment is marked, starting from a titer of 20, one unit above the map surface. The gray horizontal plane indicates the height of a titer of 50 as a reference for judging the landscapes against various estimates of neutralizing antibody correlates of protection. Additional visualizations of predicted versus fitted titers are shown in fig. S26. The number of sera included for the calculation of the landscapes are (**A**) D614G sera (n = 13), (**B**) B.1.1.7 sera (n = 13), (**C**) B.1.351 sera (n = 15), (**D**) P.1 sera (n = 13), (**E**) B.1.617.2 sera (n = 21), (**F**) B.1.526+E484K sera (n = 4), (**G**) B.1.637 sera (n = 2), (**H**) C.37 sera (n = 2), (**I**) BA.1 sera (n = 4), (**J**) BA.2 sera (n = 1), (**K**) 4-weeks post–2× mRNA-1273 sera (n = 30), (**L**) >3-months post–2× mRNA-1273 sera (n = 13), (**M**) 4-weeks post–3× mRNA-1273 sera (n = 26), (**N**) >3-months post–3× mRNA-1273 sera (n = 8), and (**O**) 4-weeks post–2× mRNA-1273.351 sera (n = 9). Interactive versions of the landscapes in each of the panels are available at https://acorg.github.io/mapping_SARS-CoV-2_antigenic_relationships_and_serological_responses.

mRNA-1273 vaccine dose and changes magnitude but maintains similar breadth at 4 weeks and >3 months after a third mRNA-1273 vaccine dose.

Molecular basis of the antigenic map topology

As shown in Fig. 3C, the locations of variants in the antigenic map point to amino acid substitutions that are shared between pre-Omicron variants with similar antigenic characteristics. For example, variants with substitutions at position 484 (E to K or Q) are positioned on the right of the map because of poorer neutralization by D614G and vaccine sera. Variants toward the top of the map [B.1.1.7, B.1.351, P.1, B.1.621 (Mu), and B.1.1.7+E484K] all have a substitution at position 501 (N to Y). Variants B.1.351 and P.1 additionally have substitutions at position 417, indicating that these changes are associated with increased reactivity to B.1.351 and P.1 sera. Variants in the lower half of the map [B.1.617.2, B.1.429 (Epsilon), C.37, and B.1.617.1 (Kappa)] all have substitutions at position 452 (L to R or Q). The Omicron variants, carrying at least 15 additional substitutions in the RBD, form a separate cluster in the lower right of the antigenic map.

To further investigate the molecular basis of these antigenic differences, we generated 10 lentivirus pseudotypes with single substitutions at positions 417, 452, 484, and 501 in different RBD contexts and measured the effect on reactivity to different serum groups and the subsequent positioning of variants in the antigenic map. As shown in Fig. 5, the antigenic effect of the different substitutions when introduced in isolation was generally consistent with that inferred from the antigenic map of wild-type variants. Large antigenic effects were seen for substitutions at position 484, which were associated with the right-left antigenic variation seen in the map, such that D614G+E484K showed greater escape than D614G+E484Q in D614G sera (Fig. 5, A and B, and fig. S27, A and B). Smaller but significant effects were seen for substitutions at position 417, which was shown to mediate some of the top-bottom map variation (Fig. 5B and fig. S28). Introduction of the N501Y substitution alone into D614G did not mediate large antigenic changes in the map but did cause significantly increased reactivity to B.1.1.7, P.1, and B.1.351 sera (fig. S29) and generated a virus that was antigenically similar to B.1.1.7, which also contains only the substitution N501Y in the RBD (Fig. 5A and fig. S1).

Despite overall correspondence with map-based predictions, some results were unexpected. In particular, although D614G+L452R showed decreased reactivity to D614G sera (fig. S27C), the D614G+L452R+E484Q mutant did not show any significant difference in reactivity compared with that of the D614G+E484Q mutant (fig. S27, B and D). Moreover, the B.1.429+K417N mutant showed increased reactivity to the D614G sera (fig. S28D), even though this represents a change away from sequence identity with D614G. Because it has been shown that the presence of the K417N substitution alone in the absence of 501Y greatly reduces angiotensin-converting enzyme 2 (ACE2) affinity in some contexts (*31*), we speculate that the effect seen in B.1.429+K417N may be influenced by an artificial inflation of titers generally.

Furthermore, the introduction of the A484K substitution into the BA.1 context allowed us to compare the effect of the alanine (A) substitution present in the Omicron variants in contrast to the lysine (K) substitution seen in pre-Omicron antigenic escape variants, such as B.1.351. We found that the BA.1+A484K mutant had lower titers compared with the BA.1 variant for sera taken 4 weeks after a second mRNA-1273 vaccine dose and also for B.1.617.2 sera (fig. S30), raising the question as to why this alternative substitution has not been seen more frequently in Omicron variants in nature. In terms of movement in the antigenic map, in two dimensions, the BA.1+A484K mutant moves to a location to the top right of the BA.1 variant, bringing it closer to the B.1.351 sera but also consequently close to BA.4/BA.5 (fig. S31). However, antigenic relations with BA.4/BA.5 are better described in 3D representation (Fig. 3B), in which the BA.4/BA.5 variant uses the third dimension and the BA.1+A484K mutant occupies a novel area of antigenic space distinct from the other Omicron variants (Fig. 5C).

Variation in immunodominance of different RBD sites between serum groups

Overall, among the substitutions tested, not all serum groups were equally sensitive to changes at a given position, with some substitutions having a large effect on reactivity to certain serum groups but little to no effect in others. For example, although the D614G+E484K and D614G+E484Q substitutions had large effects on reactivity to D614G sera, no significant effect on B.1.351 serum reactivity was found (fig. S27, A and B). Such findings are consistent with variation in immunodominance patterns and in the extent to which antibodies in different sera target different structural regions in the RBD.

We sought additional evidence for immunodominance switches by analyzing results from the mutants alongside differences in serum reactivity between other pairs of variants that differed by single–amino acid substitutions in the RBD (Fig. 6A and figs. S32 and S33).

For all pairs of variants with differences in the RBD only at position 484 from E to K or Q (Fig. 6A, rows 1 and 2), D614G sera consistently showed significantly reduced reactivity [by using the 95% highest-density interval (HDI) estimate of the pairwise difference in titers as a criterion throughout], whereas the B.1.351 sera showed little to no significant change in reactivity. Across the other serum groups, sera from individuals infected with variants with the ancestral 484E (2× mRNA-1273, D614G, B.1.1.7, and B.1.617.2) were generally sensitive to differences at position 484, whereas sera from infections with variants with 484K (B.1.351, P.1, and B.1.526+E484K) were more resistant to these substitutions, although B.1.351 did show evidence for some smaller increases in titers in some cases. The BA.1+A484K mutant also reflected this pattern of sensitivities in the serum groups against which it was titrated, with significant titer decreases in mRNA-1273 and B.1.617.2 sera (484E) and a smaller increase in reactivity to mRNA-1273.351 sera (484K) (fig. S30, S33). Of the four BA.1 sera (484A), the mean fold decrease of 2.1 against the BA.1+A484K mutant versus the BA.1 variant indicated that these sera are also somewhat sensitive to changes at the 484 position. However, this effect was far less than that seen for the D614G+E484K mutant on reactivity to post-D614G infection

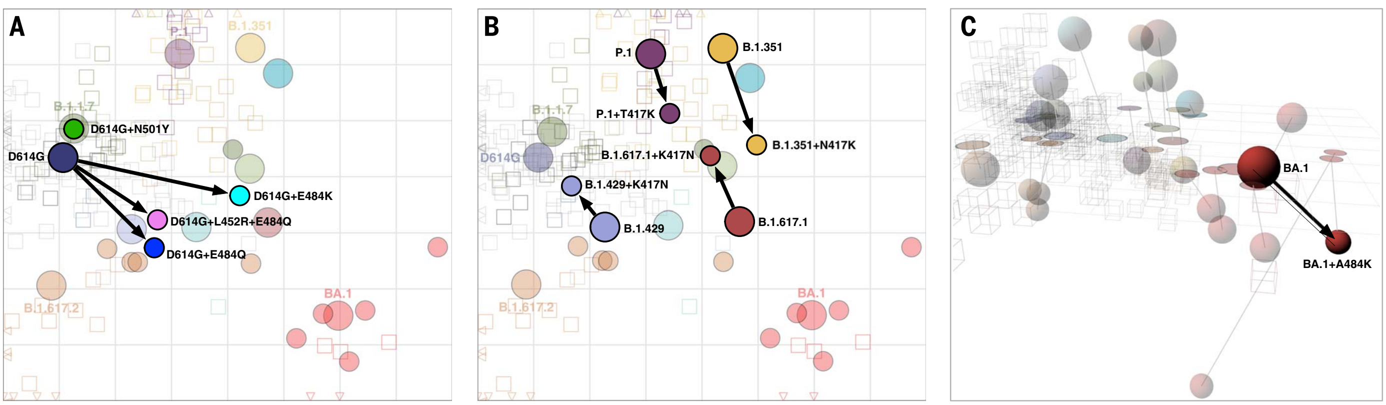


Fig. 5. Antigenic maps including laboratory-made mutants with substitutions at positions 417, 452, 484, and 501. (**A**) Variants with substitutions N501Y, E484K, E484Q, and L452R+E484K in the background of D614G; D614G+L452R is not shown because it was titrated against only D614G sera, so its position could not be determined. (**B**) Variants with the T/N417K substitution in the background of P.1 and B.1.351, respectively, and K417N in the background of B.1.429 and B.1.617.1. (**C**) BA.1 with the substitution A484K. The map in (C) is 3D to highlight the antigenic differences between BA.1, BA.1+A484K, and BA.4/BA.5. The 2D version of (C) is shown in fig. S31. Arrows point from the antigenic position of the root virus to that of the laboratory-generated variant. Interactive versions of the maps shown in each panel are available at https://acorg.github.io/mapping_SARS-CoV-2_antigenic_relationships_and_serological_responses.

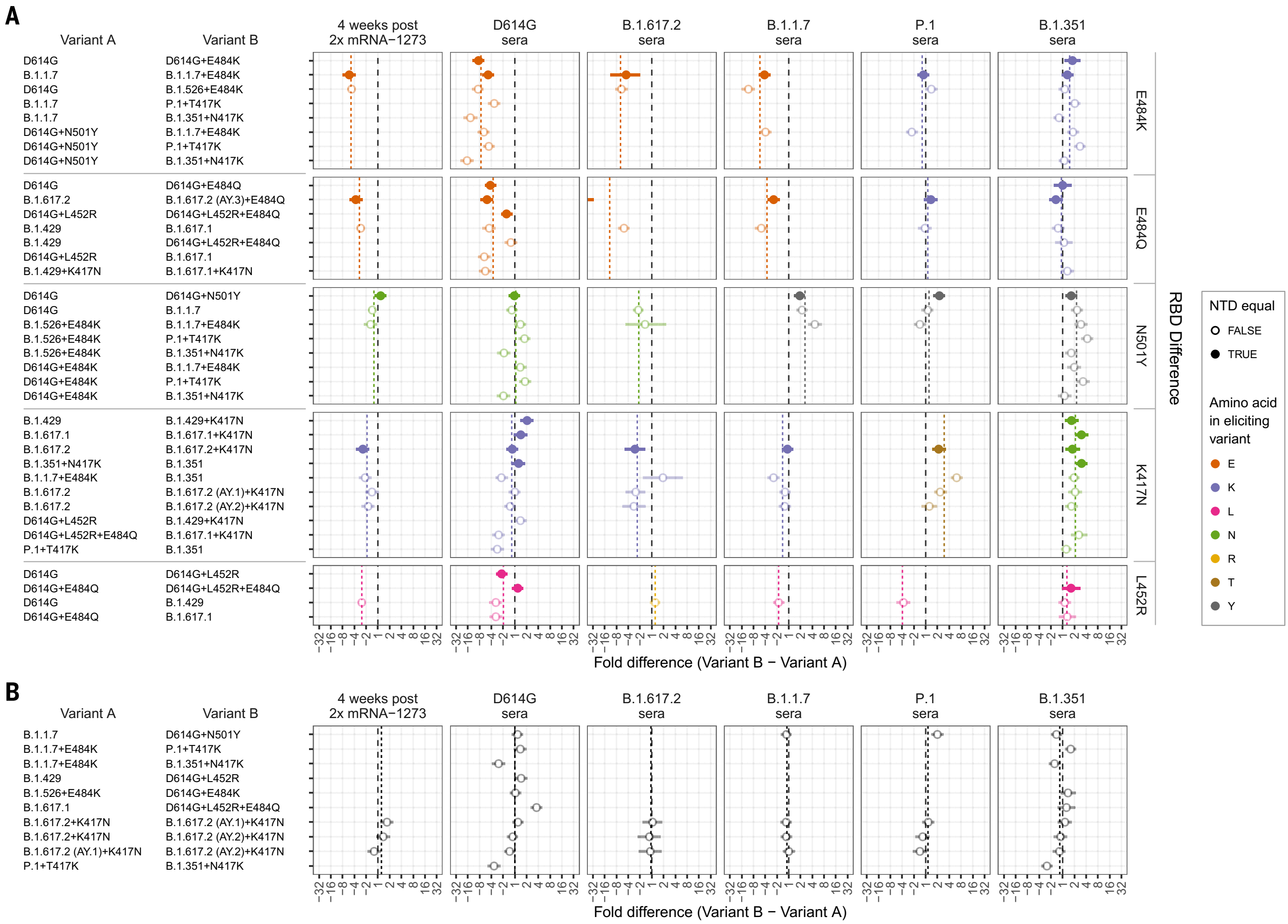


Fig. 6. Effect of pairwise amino acid differences on reactivity to different serum groups. This plot compares the average fold difference in titer between (**A**) different pairs of variants that differ by only a single amino acid difference in the RBD or (**B**) that do not differ by any amino acids in the RBD but differ in the NTD. Comparisons are grouped by serum group (panel columns) and corresponding RBD difference (panel rows). In each panel, the circle represents the estimate for the average fold difference in titer between variant A and variant B, as named on the left-hand side of the plot, and the lines extend to indicate 95% HDI for this estimate. The black dashed line marks a fold difference in titer of 1 (no difference), and the colored dashed line indicates the average fold difference between all pairs of variants with that substitution in the RBD. Points and lines are colored according to the amino acid in the variant homologous to that serum group, at the position in the RBD where the pair of variants compared differ. Filled circles indicate where pairs of variants had no additional amino differences in the NTD region, often because one was generated as an artificial mutant. By contrast, open circles indicate pairs of variants with additional amino acid differences in the NTD region, in addition to the RBD amino acid difference listed. The estimate for B.1.617.2 sera fold differences between the B.1.617.2 and B.1.617.2 (AY.3)+E484Q variants [(A), third column, second row], which falls outside the plot, is −59.4 (95% HDI −117.1, −30.7). Details of how fold-difference estimates and HDIs were calculated are described in Materials and methods. Figures S30 and S31 show the same results against an expanded set of pairwise amino acid differences. Interactive scatterplots comparing titers against each pair of variants for each serum group are available at https://acorg.github.io/mapping_SARS-CoV-2_antigenic_relationships_and_serological_responses.

and vaccination sera, suggesting that like the B.1.351 and P.1 serum responses, the BA.1 response also appeared to be directed away from the 484 position.

Where variants differed only by N501Y (Fig. 6A, row 3), we also found that sensitivity was linked to the amino acid at position 501 in the infecting variant. The two serum groups resulting from infections with B.1.351 and B.1.1.7 (both with 501Y) were sensitive to differences at position 501, whereas the serum groups postprototype vaccination and post-D614G, B.1.526+E484K, and B.1.617.2 infection (variants with ancestral 501N) were typically not sensitive to the N501Y substitution.

For K417N comparisons (Fig. 6A, row 4), there was evidence for equal or decreased titers in serum groups with exposure to variants with the ancestral K at position 417 (mRNA-1273, D614G, B.1.617.2, and B.1.1.7 serum groups) and a corresponding overall increase in titers to the serum group after infection with B.1.351 (417N). P.1 sera (417T) showed increased reactivity to the K417N substitution. The B.1.351 sera (417N) also showed increased reactivity associated with the K417T substitution, likely reflecting a structural homology of the changes caused by K417T and K417N.

In addition, where variants allowed for a comparison of the effect of the substitution L452R (Fig. 6A, row 5), we found evidence for most of the serum groups distinguishing between variants that differed by this substitution, with decreased (mRNA-1273, D614G, B.1.1.7, and P.1 sera) or increased (B.1.617.2 sera) titers, corresponding to the amino acid present at position 452 in the eliciting variant. One exception was the B.1.351 sera, in which, unlike the other serum groups with the ancestral 452L,

the L452R substitution did not produce an overall decrease in titers.

Impact of changes in the NTD

We also investigated the effects of substitutions in the N-terminal domain (NTD). Generally, the effects of single–amino acid differences in the RBD were consistent regardless of whether additional NTD differences were present or not (Fig. 6A). This was also reflected in pairwise comparisons of variants with no RBD differences, in which we typically found no significant differences in titer reactivity (Fig. 6B). This included little evidence for a significant effect on serum reactivity of features, such as the NTD 69 to 70 deletion in the B.1.1.7 variant (Fig. 6B, B.1.1.7 versus D614G+N501Y). However, for comparators involving the B.1.351+N417K mutant, which has the substantial B.1.351 NTD changes that include the 241 to 243 deletion and R246I substitution, we did find evidence of differences compared with other viruses that had sequence identity in the RBD, such as the P.1+T417K and B.1.1.7+E484K mutants. Although these differences may be associated with the effects of these NTD substitutions on titer reactivity [as has been found in other cases (*32*)], we found that in the comparisons shown in Fig. 6B, the B.1.351+N417K mutant also had lower titers against the B.1.351 sera themselves, despite sequence identity of the NTD differences for these sera. This indicates that in our data, there was a general negative bias in titers measured against the B.1.351+N417K mutant, rather than antigenic effects of the substitutions within the NTD. Others have found that sensitivity to NTD substitutions can depend on the levels of ACE2 expression in the assay target cell line (*33*), which may relate to us finding little evidence overall for NTD effects in the cases that we compared.

Discussion

The antigenic analyses of SARS-CoV-2 variants presented here underscore the advantages of an integrated and extensible framework for understanding antigenic relationships and serum responses in SARS-CoV-2. The antigenic map and the antibody landscapes allow the comparison of serum responses by variant and also provide a quantitative measurement of both the response magnitude and breadth after different exposures. Antibody landscapes can also be applied to analyze how multiexposure serum responses distribute across antigenic space (*34*), which is important for ongoing studies seeking to compare the immunity built through different prospective vaccination regimens (*35*, *36*).

Using regression analyses and antibody landscapes to analyze immune responses in vaccination sera taken at multiple time points after vaccination, we quantified and visualized changes in response cross-reactivity breadth, disentangling it from changes in cross-reactivity that are due to differences in response magnitude (Fig. 2). The similar estimates of response breadth just before the third vaccination, and from 4 weeks after the third vaccination, showed that the main short-term effect of the third vaccination was to boost the magnitude of a response that had already become more cross-reactive, rather than to generate additional breadth of cross-reactivity. This is consistent with studies into influenza vaccine responses, in which the predominant effect of vaccination is boosting of preexisting patterns of prevaccination reactivity, which was independent of the vaccine variant used (*34*). It is possible that the third vaccine dose was responsible for the later additional increases in cross-reactivity in >3 months after third vaccine dose samples, although the effect size was small and not statistically significant. It is also possible that our observation of increasing response breadth over time (Fig. 2B) would have continued to this point without a third vaccination. However, increases in breadth in the absence of a third dose would not have translated to increased protection without the effect of the third dose boosting antibody titers.

Different variants and serum responses clustered on the antigenic map in a way that allowed the inference of candidates for the amino acid substitutions responsible for the main antigenic changes. The impact of substitutions at position 484, 417, and 501 agree with other work based on deep mutational scanning (DMS) (*37*–*39*) and neutralization data from vesicular stomatitis virus pseudotyped particles (*40*, *41*). The large antigenic effect of single substitutions such as E484K is reminiscent of influenza viruses, in which a large portion of the antigenic difference is associated with only a single or double substitution located adjacent to the receptor binding site on the hemagglutinin protein (*42*–*44*). This is also observed for the determinants of SARS-CoV-2 antigenic evolution described here, which are located close to the ACE2 binding site on the spike protein (figs. S34 and S35).

Variation in immunodominance among distinct antigenic regions is gaining recognition as an important aspect of the antigenic evolution of pathogens and has also been seen in influenza, hepatitis C, and noroviruses (*45*–*47*). Conversely, the codominance of multiple sites in measles virus has been associated with reduced potential for evasion from serological reactivity (*48*). Within SARS-CoV-2, our findings are consistent with studies showing that monoclonal antibodies tend to have different regional binding preferences when sourced from individuals exposed to different variants (*10*, *49*–*51*). Conclusions suggesting an immunodominance switch for B.1.351 sera away from the 484 region also correspond with published data applying DMS to sera after D614G and B.1.351 infection (*39*). Our results show how such immunodominance switches extend to vaccination and infections with different variants and how changes can be associated with the amino acid present in the different eliciting variants. These observations explain the wider patterns in the data—in which certain serum groups may distinguish clear antigenic differences between certain variants although others may not—and again underscore that antigenic differences between variants are not necessarily absolute measures but can be relative to the particular sera against which variants were titrated and the specific structural regions that the particular sera tested predominantly target. Immunodominance switches also highlight the potential limitations of visualizing antigenic relationships solely in Euclidean space, as is done with antigenic maps, and emphasize the need for further complementary approaches to their quantification and visualization, such as the explicit fold-drop comparisons that we use here.

We note two limitations in the analyses that we have presented. First, although the assay that we use is FDA certified, it measures neutralization with a lentivirus pseudotype neutralization assay, whose results may differ from those of live-virus neutralization assays. However, we generally find our fold-difference estimates to be within the ranges reported in other studies and, in agreement with others (*12*), we do not find an obvious bias compared with live-virus neutralization assay results (fig. S7). Second, we focus on patterns of serological reactivity and antigenic variation found up until the early Omicron variants. Beyond this point, it is very difficult to source sufficient primary-exposure human sera to study antigenic relationships in detail, and it is increasingly necessary to rely on animal-model sera. Consequently, it is critical to assess whether the patterns of antigenic relatedness and changes in immunodominance found for human sera can be reliably measured in different animal models.

As populations increasingly experience multiple exposures, it will be important to investigate how antigenic differences in multiexposure serum responses compare with those inferred from primary-exposure sera, in particular with immunodominance patterns in mind. For example, findings that infection with Omicron BA.1 after a previous exposure to pre-Omicron variants predominately boosts antibodies targeting epitopes shared between pre-Omicron and Omicron variants highlight the value of understanding how different initial exposures dictate the structural regions that were initially targeted and how this interacts with a future exposure (*52*–*55*). Answers to these questions will better reveal the variants and substitutions to which different populations are most vulnerable and help anticipate which emerging variants may be most at risk of evading

current immunity. The choice of vaccine immunogens on the basis of immunodominance considerations may become as important as their antigenic characteristics.

Materials and methods

Specimens and study cohorts

All samples, including those from vaccine recipients, were collected in clinical studies with information listed in table S1. Convalescent sera are assumed but not confirmed first infections, based on no previously recorded or self-reported SARS-CoV-2 vaccination or infection, plus additional criteria in some cases, as also described in table S1. For convalescent samples, the infecting variant was determined using whole-genome sequencing. Institutional review board (IRB) approvals were obtained for each study site. All participants were given and signed the informed consent form.

Vaccination samples came from a variety of sources, as also indicated in table S1 and described in more detail below:

4 weeks post 2× mRNA-1273

The majority (n = 28) of the 4-weeks post–2× mRNA-1273 samples were from the mRNA-1273 phase 1 study (NCT04283461) and were obtained from the Division of Microbiology and Infectious Diseases (DMID), National Institute of Allergy and Infectious Diseases (NIAID) for the mRNA-1273 phase 1 study team, and Moderna. The phase 1 study protocols and results are reported previously (57, 58). The phase 1 trial tested the identical vaccine (mRNA-1273), dose (100 μg), and schedule as used in the Moderna phase 3 (NCT04470427); titrations against these samples were also reported in (59). A smaller subset of samples (n = 4) were obtained from the Coronavirus Efficacy (COVE) phase 3 trial and were preselected as having high titers of neutralizing antibodies against D614G to enable detection of neutralization against the wider range of Omicron variants that they were measured against; some titrations against these samples also appear in (60). Finally, an additional 11 sera from the COVE phase 3 trial were titrated against D614G and the D614G+N501Y mutant, and an additional 26 sera were titrated against BA.1 and the BA.1+A484K mutant. These sera appear in the results shown in fig. S29 and S30, respectively, but are not included in the general analysis due to lack of titrations against further variants.

>3 months post 2× mRNA-1273

All samples relate to sera previously described in (61) that were obtained before homologous mRNA-1273 boosting (50 μg) in 16 study participants who had received two inoculations of mRNA-1273 (100 μg) under emergency-use authorization. These participants were part of a subset analysis of high D614G responders in the NIAID heterologous boost study DMID 21-0012.

4 weeks post 3× mRNA-1273

Sixteen samples relate to sera from the same individuals in the >3-months post 2× mRNA-1273 group, but obtained 29 days after homologous mRNA-1273 boosting (50 μg), again as described and appearing in (61). An additional 10 samples also relate to samples collected as part of the DMID 21-0012 and also relate to samples previously described in (60), which were also a subset of high D614G responders.

>3 months post 3× mRNA-1273

These samples were taken as part of the COVID-19 Variant Immunologic Landscape (COVAIL) randomized clinical trial (NCT 05289037) and are from the subset of individuals who had received a previous mRNA-1273 vaccine regimen followed by a mRNA-1273 booster dose. Samples were taken directly prior to a fourth mRNA-1273 dose and, as for other vaccine groups, we subset to those with no previous record of a SARS-CoV-2 infection.

4 weeks post 2× mRNA-1273.351

Samples were from COVID-19 naïve participants who received two 100-μg doses of mRNA-1273.351 on days 1 and 29 as part of a clinical trial (NCT 04785144), experimental arm 2C.

The use of the specimens for the research conducted in this manuscript is covered under IRB protocols Pro00093087 and Pro00105358.

Where samples in the same serum group came from a mixture of cohorts or agencies, we did not find evidence for differences in the patterns of reactivity according to cohort or agency (fig. S3).

Pseudotype virus assay

Neutralization was measured with lentiviral particles pseudotyped with SARS-CoV-2 spike and containing a firefly luciferase (Luc) reporter gene for quantitative measurements of infection by relative luminescence units (RLU). The assay was performed in 293T/ACE2-MF provided by Michael Farzan (-MF) and Huihui Mu. The derivation of the cell line is described in (62). Pseudoviruses were prepared, titrated, and used for measurements of neutralizing antibodies essentially as described previously (59). Briefly, an expression plasmid–encoding codon-optimized full-length spike of the Wuhan-1 ancestral sequence (VRC7480) was provided by Barney Graham and Kizzmekia Corbett at the Vaccine Research Center, National Institutes of Health (USA). Mutations were introduced into VRC7480 either by site-directed mutagenesis using the QuikChange Lightning Site-Directed Mutagenesis Kit from Agilent Technologies (Catalog no. 210518) or were created with a spike gene synthesized by GenScript, using the spike sequence in VRC7480 as template. All mutations were confirmed with full-length spike-gene sequencing by Sanger Sequencing, using Sequencher and SnapGene for sequence analyses. Pseudovirions were produced in HEK293T/17 cells (ATCC cat. no. CRL-11268) by transfection using Fugene 6 (Promega Cat no. E2692) and a combination of spike plasmid, lentiviral backbone plasmid (pCMV ΔR8.2), and firefly Luc reporter gene plasmid (pHR′ CMV Luc)36 in a 1:17:17 ratio in Opti-MEM (Life Technologies) (CMV, cytomegalovirus). Transfection mixtures were added to preseeded HEK 293T/17 cells in T-75 flasks containing 12 ml of growth medium and incubated for 16 to 20 hours at 37°C. Medium was removed and 15 ml of fresh growth medium added. Pseudovirus-containing culture medium was collected after an additional 2 days of incubation and clarified of cells by low-speed centrifugation and 0.45-μm micron filtration. Median tissue culture infectious dose ($TCID_{50}$) assays were performed as described previously (59).

For measurements of neutralization, a pretitrated dose of virus was incubated with eight serial fivefold dilutions of serum samples (1:10 or 1:20 starting dilution) in duplicate in a total volume of 150 μl for 1 hour at 37°C in 96-well flat-bottom culture plates. 293T/ACE2-MF cells were detached from T-75 culture flasks using TrypLE Select Enzyme solution, suspended in growth medium (100,000 cells/ml) and immediately added to all wells (10,000 cells in 100 μl of growth medium per well). One set of eight wells received cells and virus (virus control), and another set of eight wells received cells only (background control). After 66 to 72 hours of incubation, medium was removed by gentle aspiration, and 30 μl of Promega 1X lysis buffer was added to all wells. After a 10-min incubation at room temperature, 100 μl of Bright-Glo luciferase reagent was added to all wells. After 1 to 2 min, 110 μl of the cell lysate was transferred to a black or white plate. Luminescence was measured using a GloMax Navigator luminometer (Promega). Serum samples were heat-inactivated for 30 min at 56°C prior to assay. Neutralization titers are the inhibitory dilution (ID) of serum samples at which RLUs were reduced by either 50% (ID_{50}) or 80% (ID_{80}) compared with virus control wells after subtraction of background RLUs. The spike mutations of all variants used in this study are shown in table S2.

Titer analyses

Excluding outlier sera

The serum samples used in this study were collected from individuals with no reported previous vaccination or infection preceding the exposure that was associated with their respective serum group, with further additional criteria for some sample sources as described in table S1. However, as with most studies, it is possible that for some individuals, additional previous (even asymptomatic)

infections were not detected or reported. Based on this possibility, we sought to identify sera with patterns of reactivity that were less likely to be consistent with a true first exposure. As such, we applied selection criteria to all the sera to identify cases where a titer was measured that was higher than the serum-group-homologous-variant titer by more than twofold (for example, a serum from the B.1.617.2 group that had a titer against D614G that was more than twofold higher than the titer measured against B.1.617.2). Subvariants of B.1.617.2 and BA.1 (Fig. 3, smaller circles) were excluded from this analysis, as well as the variant P.1, which we noted tended to often have high titers generally. For example, P.1 often has higher-than-homologous titers for sera from the B.1.351 serum group, possibly due to assay-related biases inflating titers for the particular P.1 variant, as it has been observed in other assays using lentiviral pseudotypes (9). B.1.429 was used as the homologous variant for sera from the B.1.637 group, being the genetically most similar variant titrated, while D614G was used for mRNA-1273 vaccination groups; and B.1.351 was used for the mRNA-1273.351 vaccination group. Fifteen of the total of 207 sera had titers exceeding the twofold-higher-than-homologous criteria described above and are highlighted in red in the titer plot shown in fig. S2.

As an additional step, we also excluded nine sera as likely second-infection responses based on visual inspection of patterns of reactivity that deviated strongly from the general trend of titer reactivity for a given serum group. These sera are additionally highlighted in fig. S2 in blue. In particular, this criteria applied to sera from the BA.1 serum group where we noted a clearly bimodal pattern of reactivity with three post-BA.1 infection sera that showed high titers against the D614G variant and four sera which did not. Based on the large antigenic difference between D614G and BA.1 and the types of reactivity seen in more controlled animal-model BA.1 infections (63), we judged the four more-BA.1-specific sera to be likely first infections and excluded the rest of the group as likely cases that had had a previous exposure to a pre-Omicron variant.

Calculation of Geometric Mean Titers (GMT) and fold changes for nondetectable titers

Calculation of the mean titers becomes nontrivial when experiments contain nondetectable titers or, in data science terminology, censored data. Although computation of means of data with censored values has been around for over 60 years (64–66), it does not seem to be commonly employed when analyzing data from neutralization assays. Correct estimation of the GMT becomes especially important if one wants to quantify the amount of escape by variants like the Omicron variants, with many nondetectable titers. Setting nondetectable titers to half the value of an assay's limit of detection is a common method employed in virology, however, it may overestimate the GMTs for such variants. Below we give a brief overview of our approach to dealing with such censored data, functions for which are included in the associated titertools package (67).

The main step in estimation of the titers is the construction of a likelihood function that describes the probability of observing both the nondetectable and detectable titers. In the dataset used for this study, we deal with only left-censored data. We assume that $\log_2$ titer values of a particular serum group are modeled by a normal density function, $f(x;\mu,\sigma)$, with unknown mean μ and standard deviation (SD) σ, and we let $y_i(i = 1,...,M)$ denote detectable titers and $T_i(i = 1,...,N)$ denote the nondetectable titers. Then the log likelihood function, LL, is given by

$$\mathrm{LL}(\mu,\sigma) = \sum_{i=1}^{N} \log\int_{-\infty}^{T_i} f(x;\mu,\sigma) + \sum_{i=1}^{M} \log f(y_i;\mu,\sigma)$$

In this expression, the term involving the integral is the likelihood of observing the censored titers, whereas the other term is the likelihood of observing the noncensored titers. This likelihood function then can either be used in a maximum-likelihood estimation pipeline or, if priors are chosen for μ and σ, in a Bayesian posterior-calculation pipeline from which the value of μ can be estimated in various ways. For the calculations in this paper, confidence intervals were calculated using the highest density interval (HDI) method of calculation from 10,000 samples.

When calculating the fold changes in titers, the same approach to nondetectable titers was taken, with, for example, a change from 40 to <20 equating to a change of <–1 on the $\log_2$ scale, for which the likelihood function would be the integral of the normal density function from $-\infty$ to –1. For estimating fold changes, we used the function "log2diff" in the titertools package, with parameter "sigma = 0.87," based on our estimates of the magnitude of variation in paired measurement repeats (fig. S10).

Titers after accounting for estimated individual effects

The data used in this paper contain sera collected from different individuals infected with the same variants. Even when these titers are grouped according to the infecting variant, one can see quite large variations in titer magnitudes, whereas the qualitative appearance of the titer curve is similar (Fig. 1). Some of this is due to individual variation in the overall reactivity of sera collected from different individuals, for example, related to the magnitude of the serological response. We model this behavior in the following way: Assume that t_{ij} is the $\log_2$ titer measurement of antigen i against serum j from the serum group J. Then

$$t_{ij} = s_{iJ} + r_j + \varepsilon_{ij} (i = 1, \ldots n \text{ and } j = 1, \ldots m)$$

where s_{iJ} is the average $\log_2$ titer of antigen i against the serum group J, r_j is the serum reactivity bias of the serum j, and ε_{ij} is the independently and normally distributed log-titer noise. n is the number of antigens, and m is the number of sera. For each serum group, parameters s_{iJ} and r_j are then found to maximize the sum of the likelihood of each measured log titer given the observed log titer, t_{ij}. For the fitting procedure, we used the function "estimate_sr_effects" in the titertools package, with parameter "sigma = 0.62," according to our estimates of the magnitude of measurement noise in a single measurement (fig. S10).

Antigenic cartography has its own method of dealing with serum reactivity bias. The formula for the target distance (provided in "Antigenic cartography" below) indicates that for any serum, the antigen with the highest titer (usually the homologous antigen, but not always) is automatically adjusted to have 0 distance to that serum. This is true regardless of the value of the maximum titer, and this adjustment also takes care of the individual serum reactivity in a way that is compatible with antigenic mapmaking.

Calculating fold-drop differences in vaccine sera

When exploring differences between postinfection and different postvaccination responses homologous to the same antigen, one might expect that the fold drop in serum responses (compared with the homologous antigen) are similar in some way. A simple model one can use is to introduce an overall factor that relates the responses of sera from the two groups. In terms of the model in the previous section, this means

$$s_{hJ} - s_{iJ} = d_J(s_{hK} - s_{iK}) = d_J \Delta s_{hi}{}^K$$

where K and J are two serum groups, h is the homologous antigen to them, i is any other antigen, and s_{xY} are as described in the previous section. Note that terms such as $s_{hJ} - s_{iJ}$ give the (noiseless) fold drop of titers of antigen i from the titers of the antigen h (homologous to J), measured against the serum group J. Then for a specific serum j from the serum group J, we have

$$\begin{aligned} t_{hj} - t_{ij} &= s_{hJ} - s_{iJ} + \varepsilon_{hj} - \varepsilon_{ij} \\ &= d_J \Delta s_{iK} + \varepsilon_{ijh} \end{aligned}$$

where the last term is again normally distributed however with $\sqrt{2}$ times the SD of the titer noise.

An example would be K = D614G postinfection serum group, J = mRNA-1273 serum group, h = D614G antigen, and i = any other antigen (Fig. 2). The term d_J is the "factor" specific to serum group J, which relates its responses to the response of serum group K. (This term obviously also depends on K; however, we imagine that K is a constant postinfection serum group and that J varies between different postvaccination serum groups, so by abuse of notation, we only index this term with J.) Note that taking difference of titers cancels out the serum response bias specific to serum j.

When applying this model, we compared the serum groups D614G, 4 weeks post 2× mRNA-1273, >3 months post 2× mRNA-1273, and 4 weeks post 3× mRNA-1273 (with shared homologous variant D614G), and we compared serum groups B.1.351 and 2× mRNA-1273.351 (with shared homologous variant B.1.351). In the D614G and B.1.351 postinfection groups, we fixed d_J as 1, estimating fold-change differences in the post-vaccination groups retive to the postinfection sera. For the fitting procedure, we used the function "compare_crossreactivity_breadth" in the titertools package, with confidence intervals calculated for the parameters based on the highest density interval of 5000 sampling runs after 1000 warmup runs.

Imputing censored titers

For figs. S11, S12, and S20, we imputed nondetectable titers in order to not bias the distributions that are visualized. In these cases, censored titers (e.g., <20) are imputed using the function "impute_gmt_titers" from the titertools package. In these cases, censored titers are replaced with numeric titers drawn from a censored normal distribution with mean and SD parameters that are estimated from the titer data as described in the "Calculation of GMT" section above.

Antigenic cartography

Antigenic cartography is a multidimensional scaling method (MDS) (68) for the reduction of dimensions in large-titer datasets in order to not only visualize large amounts of titer data in a lower-dimensional space but to also draw conclusions about the antigenic evolution of pathogens by studying the positions occupied by these viruses on the antigenic map.

In order to construct an antigenic map, one starts with a titer table where columns are sera and rows are antigens. Denoting the values of such a table by t_{ij} (where i denotes the antigen index and j the sera index), titer values represent a degree of dissimilarity of the ith antigen to the jth serum. We turn this dissimilarity score into a similarity score by the following transformation:

$$D_{ij} = \max_i [\log_2(t_{ij})] - \log_2(t_{ij})$$

The D_{ij} values are called the "target distances." Antigenic cartography strives to find a multidimensional (usually two or three dimensions) placement of antigens and sera, such that the Euclidean distance between antigens and sera conform as well as possible to the target distances. This is achieved by an optimization of the antigen and serum coordinates that minimize the stress function

$$\sum_{i,j} (D_{ij} - d_{ij})^2$$

where d_{ij} represents the Euclidean distance between the ith antigen and the jth serum points in the antigenic map. The minimization is achieved using the L-BFGS algorithm (69) to find the best cost-minimizing coordinates starting from a set of randomly determined initial conditions. Note unlike most common MDS problems, the similarity matrix is not square and might contain masked or censored values and therefore requires a slightly custom approach. In particular, if t_{ij} is a nondetectable titer, then the stress is computed as

$$\varphi(D_{ij} - d_{ij} + s)(D_{ij} - d_{ij} + s)^2$$

where s is a step-size factor which is 1 for discrete data (and their means) such as those obtained from Haemagglutinin Inhibition assays, or 0 when the output data are continuous and measured from titer curves, and

$$\varphi(x) = \frac{1}{1 + 10^{-x}}$$

is a damping function whose role is to give more freedom to nondetectable titers to be optimized for a value less than the limit of detection of the assay.

We used the Racmacs package to compute the antigenic maps presented in this manuscript (70). The map was constructed using 1000 optimizations, dilution step size of 0, with the minimum column basis parameter set to "none." In order to allow for accurate coordination of serum positions in two dimensions, sera with less than three detectable titers were excluded from the maps. An exception was made for the BA.2 serum sample, which was included despite only having two detectable titers because it was the only serum of its group, although the serum position for this serum will therefore be less constrained in its antigenic position.

Assessing model fit

We performed different analyses to ascertain that the 2D antigenic map is a good representation of the titer data and to ensure that the map is robust to noise from measurement errors.

First, to get an understanding of the measurement error of individual titrations, we analyzed a set of repeat titrations of 60 sera against eight of the variants used in the assay and found that the mean difference in $\log_2$ titers between repeated titrations was 0.209 with a SD of 0.84, with measurement error being approximately normally distributed (fig. S10). Accounting for the systematic bias of titrations between repeats, and assuming a mean of 0 without it, resulted in a total SD of 0.87 on the $\log_2$ scale. Since measurement error would be present for both the first and second repeat titration, this would be consistent with noise per measurement with a SD of $\sqrt{0.87^2/2} = 0.62$.

Subsequently, we investigated how well antigenic maps of different dimensions represent the measured patterns of serological reactivity. In our previous experience with other antigenically variable pathogens, the dimensionality of relationships can be lower [for example, two dimensions in seasonal influenza and dengue virus (13, 20)] or higher [for example, three dimensions for A/H5N1 and swine influenza (17)] or may simply not be well represented in Euclidean space. Under cross-validation, a 2D map represented the data best (fig. S9A). The overall arrangement of the variants and sera in three dimensions is similar to that in two dimensions (fig. S9, B andC), with the exception of the BA.4/BA.5 variant, which in three dimensions occupies a position more distant from BA.1 and closer to B.1.617.2 and B.1.351 than visualized in two dimensions. The dimensional analysis was performed with Racmacs by using the function "dimensionTestMap."

Assessing goodness of fit using fitted and measured titers

Next, we assessed how well the 2D map fits the measured titers. We first investigated whether the measured $\log_2$ titers correlated with the titers fitted in the antigenic map. Overall, the fitted and measured titers corresponded well (fig. S11). When comparing the mean differences of the fitted versus measured titers (mean of 0.04 and SD of 1.05, shown as a histogram in fig. S12), they are comparable to the estimated size of measurement error (fig. S10). This suggests that most of the residual error observed in the antigenic map fit of the data are consistent with measurement error but that not all variation in the titers is captured.

We investigated whether any variant and serum group combinations exhibited consistently larger differences between fitted and measured titers (fig. S13). We find some variants showing differences in fitted versus measured titers in some serum groups, such as the B.1.617.2 variant having larger measured than fitted titers against the B.1.617.2 sera. As discussed in the main text, we speculate that when compared with other datasets (fig. S7), our measurements of B.1.617.2 serum antibodies against B.1.617.2 were biased toward higher values. Also as described in the main text, in cases such as the BA.4/BA.5 variant, a third dimension is required to better reflect the measured patterns

of reactivity. However, some of these systematic residual errors in the map fit will also be reflective of the fact that not all patterns of cross-reactivity between the serum groups can be captured in the constraints of the Euclidean space of an antigenic map. In particular, some limitations may relate to the immunodominance differences identified in this paper, whereby one serum may see an antigenic difference between two variants and one serum may not. Antigenic cartography allows for this to some extent by permitting serum positions not to be coincident with homologous antigen positions. This means that a serum point can be positioned directly between two variants for which it sees no difference, thus being antigenically equidistant from both. However, this cannot be reconciled in cases where a serum has maximum reactivity to two variants and hence an expected antigenic distance of 0 between them. Overall though, we judge that the 2D map is a reasonable representation of the broader scale underlying antigenic relationships among variants, providing a useful complement to raw titer analysis, and in some cases may compensate for systematic titer biases like that of P.1 that are otherwise spurious.

Assessing the robustness of variant and serum positions to measurement error and presence or absence of variants and serum groups

We performed three experiments to determine the uncertainty of the positions of variants and sera in the antigenic map. First, we estimated how titration measurement error and variability in antigen reactivity would affect conclusions about antigenic map positions (fig. S14). We performed a "smooth" bootstrap with normally distributed noise added to the titer measurements and/or to antigen reactivity. Noise added to the titer measurement had a SD of 0.62, in keeping with the SD of repeat titrations shown in fig. S10, and random noise for antigen reactivity had a SD of 0.4, indicating that systematic biases in titers like that suspected for P.1 may also be present to some degree for other strains in the assay. We performed 500 repeats of adding random noise (using the Racmacs function bootstrapMap, method = "noisy") and summarized the variant positions by showing the area in which 68% (one SD) of the positional variation of a serum or antigen is captured. In general, we found that the antigen and serum positions are robust to both the random and systematic noise added, without affecting the main conclusions about the antigenic relationships between the variants included.

Next, we assessed whether the positions of the variants are robust to the exclusion of each variant in turn. We find that positions of variants are robust to the exclusion of variants (fig. S15), with the small movement observed with the exclusion of D614G and B.1.351. We then investigated whether positions of variants are robust to the exclusion of each serum group in turn (fig. S16). Overall, the map was robust to the exclusion of single serum groups, with the greatest movement observed when excluding the 2× mRNA-1273.351 vaccine serum group and the B.1.617.2 and BA.1 convalescent serum groups. Relatedly, since we incorporated the information from the different post–mRNA-1273 vaccine serum groups into the map, we also explored how the antigenic map compared if constructed from postinfection sera alone (fig. S17). For some variants that had not been titrated against post-D614G infection sera (and were thus less well coordinated without including post-D614G vaccination sera titrations), we did see larger discrepencies in antigenic position in the postinfection sera–only map, in particular for the BA.4/BA.5 variant. (fig. S17A). For other variants, antigenic relationships were consistently well described in the presence or absence of postvaccination sera (fig. S17B). We also included the outlier sera identified in fig. S2 and found little difference in map topology (fig. S19).

We also evaluated the robustness of map positions on particular titrations by performing 500 "Bayesian bootstrap" repeats, where in each repeat, titers were reweighted according to a dirichlet distribution. This method is statistically similar to the more common method of bootstrapping using random sampling with replacement, but it ensures that titers are never fully removed from the dataset, which can happen when sampling with replacement and lead to underconstrained antigenic map solutions in some of the bootstrap repeats. We performed this reweighting on both variants and sera combined, as well as antigens and sera separately, using the Racmacs function bootstrapMap, method = "bayesian" (fig. S20). Map positions were largely robust to this reweighted sampling of titers.

We also show that the error is approximately equally distributed between all variants and sera (fig. S21A). Positions of variants and sera are also robust when allowing the variant and sera to move in the map while not increasing the overall map root mean square error (RMSE) between map and table distance by more than one unit (fig. S21B).

Assessing predictive power of the antigenic map (cross-validation)

To investigate whether the antigenic map is able to predict missing titers, we performed 500 cross-validation replicates, where in each replicate we removed 10% of all titrations, and predicted the missing titers from the resulting map. On the $\log_2$ scale, where difference of 1 represents a predicted titer that is 2-fold higher than the measured titer and −1 represents a titer that is twofold lower, the mean difference between predicted and measured titer was 0.53 with a SD of 1.61 (fig. S22). We also considered the difference between predicted and measured titers split by variant and serum group (fig. S23). As for figs. S12 and S13, these results imply that not all patterns of cross-reactivity between the serum groups can be captured in the constraints of the 2D Euclidean space of the map but that overall, the 2D relationships are largely predictive of titers among sera and variants.

Construction of the antibody landscapes

The construction of the antibody landscapes differs from the methodology used in previous publications (*34*, *71*) , since each serum represents a likely first infection and was used to construct the map itself with explicit assumptions about how reactivity should be distributed. Given a serum with coordinates (x_m, y_m), let t_m be its maximum $\log_2$ titer/10. Then (x_m, y_m, t_m) represents the highest point of the antibody landscape for this serum. Given any other point $p = (x, y)$ on the 2D base space, the third coordinate, t, of the surface at this point is given by

$$t = t_m - d(p, p_m)$$

where $d(p, p_m)$ is the Euclidean distance between p and p_m. In particular, the serum point and maximum titer for a given serum describe how reactivity would be expected to vary in a cone-like fashion across antigenic space, with its apex at the serum point at a height equal to the maximum measured $\log_2$ titer/10 and diminishing at a rate of a twofold reduction in titer with each antigenic unit on the map. The serum group averages shown in Fig. 4 therefore simply represent an average of all these individual slope = 1 cone-like responses for the relevant sera, displaying the GMT of all the individual titer predictions for the group.

In this case, since samples were collected after a presumed single exposure, individual landscapes simply visualize the slope 1 cone-like model of reactivity that is fit when constructing the antigenic map, rather than refitting the same data with a nonparametric modeling approach that can also be applied.

REFERENCES AND NOTES

1. World Health Organization, Weekly epidemiological update on COVID-19 - 25 May 2023; https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19—25-may-2023.
2. B. Korber *et al*., Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812–827.e19 (2020). doi: 10.1016/j.cell.2020.06.043; pmid: 32697968
3. Y. J. Hou *et al*., SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **370**, 1464–1468 (2020). doi: 10.1126/science.abe8499; pmid: 33184236
4. D. Weissman *et al*., D614G Spike Mutation Increases SARS CoV-2 Susceptibility to Neutralization. *Cell Host Microbe* **29**, 23–31.e4 (2021). doi: 10.1016/j.chom.2020.11.012; pmid: 33306985

5. World Health Organization, "Tracking SARS-CoV-2 variants," last updated 17 August 2023; https://www.who.int/en/activities/tracking-SARS-CoV-2-variants.
6. A. Z. Mykytyn *et al*., Antigenic mapping of emerging SARS-CoV-2 omicron variants BM. 1.1. 1, BQ. 1.1, and XBB. 1. *Lancet Microbe* **4**, E295–E295 (2023). doi: 10.1016/S2666-5247(22)00384-6; pmid: 36657480
7. W. F. Garcia-Beltran *et al*., Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* **184**, 2523 (2021). doi: 10.1016/j.cell.2021.04.006; pmid: 33930298
8. C. Lucas *et al*., Impact of circulating SARS-CoV-2 variants on mRNA vaccine-induced immunity. *Nature* **600**, 523–529 (2021). doi: 10.1038/s41586-021-04085-y; pmid: 34634791
9. L. Dupont *et al*., Neutralizing antibody activity in convalescent sera from infection in humans with SARS-CoV-2 and variants of concern. *Nat. Microbiol.* **6**, 1433–1442 (2021). doi: 10.1038/s41564-021-00974-0; pmid: 34654917
10. C. Liu *et al*., Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and convalescent serum. *Cell* **184**, 4220–4236.e13 (2021). doi: 10.1016/j.cell.2021.06.020; pmid: 34242578
11. W. Dejnirattisai *et al*., SARS-CoV-2 Omicron-B.1.1.529 leads to widespread escape from neutralizing antibody responses. *Cell* **185**, 467–484.e15 (2022). doi: 10.1016/j.cell.2021.12.046; pmid: 35081335
12. A. Netzl, S. Tureli, E. LeGresley, B. Mühlemann, S. H. Wilks, D. J. Smith, Analysis of SARS-CoV-2 Omicron Neutralization Data up to 2021-12-22. bioRxiv 2021.12.31.474032 [Preprint] (2022). doi: 10.1101/2021.12.31.474032
13. D. J. Smith *et al*., Mapping the antigenic and genetic evolution of influenza virus. *Science* **305**, 371–376 (2004). doi: 10.1126/science.1097211; pmid: 15218094
14. Y. Wang, I. Davidson, R. Fouchier, E. Spackman, Antigenic Cartography of H9 Avian Influenza Virus and Its Application to Vaccine Selection. *Avian Dis.* **60** (Suppl), 218–225 (2016). doi: 10.1637/11087-041015-Reg; pmid: 27309058
15. A. L. Woodward *et al*., Development of a surveillance scheme for equine influenza in the UK and characterisation of viruses isolated in Europe, Dubai and the USA from 2010-2012. *Vet. Microbiol.* **169**, 113–127 (2014). doi: 10.1016/j.vetmic.2013.11.039; pmid: 24480583
16. C. S. Anderson, P. R. McCall, H. A. Stern, H. Yang, D. J. Topham, Antigenic cartography of H1N1 influenza viruses using sequence-based antigenic distance calculation. *BMC Bioinformatics* **19**, 51 (2018). doi: 10.1186/s12859-018-2042-4; pmid: 29433425
17. S. Chepkwony *et al*., Genetic and antigenic evolution of H1 swine influenza A viruses isolated in Belgium and the Netherlands from 2014 through 2019. *Sci. Rep.* **11**, 11276 (2021). doi: 10.1038/s41598-021-90512-z; pmid: 34050216
18. D. M. Danilenko *et al*., Antigenic and Genetic Characterization of Swine Influenza Viruses Identified in the European Region of Russia, 2014-2020. *Front. Microbiol.* **12**, 662028 (2021). doi: 10.3389/fmicb.2021.662028; pmid: 33936018
19. A. Lorusso *et al*., Antigenic relationship among zoonotic flaviviruses from Italy. *Infect. Genet. Evol.* **68**, 91–97 (2019). doi: 10.1016/j.meegid.2018.11.023; pmid: 30517880
20. L. C. Katzelnick *et al*., Dengue viruses cluster antigenically but not as discrete serotypes. *Science* **349**, 1338–1343 (2015). doi: 10.1126/science.aac5017; pmid: 26383952
21. L. C. Katzelnick *et al*., Antigenic evolution of dengue viruses over 20 years. *Science* **374**, 999–1004 (2021). doi: 10.1126/science.abk0058; pmid: 34793238
22. T. Nolden *et al*., Comparative studies on the genetic, antigenic and pathogenic characteristics of Bokeloh bat lyssavirus. *J. Gen. Virol.* **95**, 1647–1653 (2014). doi: 10.1099/vir.0.065953-0; pmid: 24828330
23. A. B. Ludi *et al*., Antigenic variation of foot-and-mouth disease virus serotype A. *J. Gen. Virol.* **95**, 384–392 (2014). doi: 10.1099/vir.0.057521-0; pmid: 24187014
24. B. Ying *et al*., Protective activity of mRNA vaccines against ancestral and variant SARS-CoV-2 strains. *Sci. Transl. Med.* **14**, eabm3302 (2022). doi: 10.1126/scitranslmed.abm3302; pmid: 34846168
25. K. Röltgen, S. D. Boyd, Antibody and B cell responses to SARS-CoV-2 infection and vaccination. *Cell Host Microbe* **29**, 1063–1075 (2021). doi: 10.1016/j.chom.2021.06.009; pmid: 34174992
26. Z. Wang *et al*., Naturally enhanced neutralizing breadth against SARS-CoV-2 one year after infection. *Nature* **595**, 426–431 (2021). doi: 10.1038/s41586-021-03696-9; pmid: 34126625
27. E. M. Anderson *et al*., SARS-CoV-2 infections elicit higher levels of original antigenic sin antibodies compared with SARS-CoV-2 mRNA vaccinations. *Cell Rep.* **41**, 111496 (2022). doi: 10.1016/j.celrep.2022.111496; pmid: 36261003
28. K. L. Lynch *et al*., Magnitude and Kinetics of Anti-Severe Acute Respiratory Syndrome Coronavirus 2 Antibody Responses and Their Relationship to Disease Severity. *Clin. Infect. Dis.* **72**, 301–308 (2021). doi: 10.1093/cid/ciaa979; pmid: 33501951
29. W. Wang *et al*., Antigenic cartography of well-characterized human sera shows SARS-CoV-2 neutralization differences based on infection and vaccination history. *Cell Host Microbe* **30**, 1745–1758.e7 (2022). doi: 10.1016/j.chom.2022.10.012; pmid: 36356586
30. A. Rössler *et al*., BA.2 omicron differs immunologically from both BA.1 omicron and pre-omicron variants. *Nat. Commun.* **13**, 7701 (2022). .doi: 10.1038/s41467-022-35312-3
31. M. Yuan *et al*., Structural and functional ramifications of antigenic drift in recent SARS-CoV-2 variants. *Science* **373**, 818–823 (2021). doi: 10.1126/science.abh1139; pmid: 34016740
32. I. Kimura *et al*., The SARS-CoV-2 Lambda variant exhibits enhanced infectivity and immune resistance. *Cell Rep.* **38**, 110218 (2022). doi: 10.1016/j.celrep.2021.110218; pmid: 34968415
33. A. G. Farrell *et al*., Receptor-Binding Domain (RBD) Antibodies Contribute More to SARS-CoV-2 Neutralization When Target Cells Express High Levels of ACE2. *Viruses* **14**, 2061 (2022). doi: 10.3390/v14092061; pmid: 36146867
34. J. M. Fonville *et al*., Antibody landscapes after influenza virus infection or vaccination. *Science* **346**, 996–1000 (2014). doi: 10.1126/science.1256427; pmid: 25414313
35. S. Chalkias *et al*., A Bivalent Omicron-Containing Booster Vaccine against Covid-19. *N. Engl. J. Med.* **387**, 1279–1291 (2022). doi: 10.1056/NEJMoa2208343; pmid: 36112399
36. A. R. Branche *et al*., SARS-CoV-2 Variant Vaccine Boosters Trial: Preliminary Analyses. medRxiv 2022.07.12.22277336 [Preprint] (2022). .doi: 10.1101/2022.07.12.22277336
37. A. J. Greaney *et al*., Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. *Nat. Commun.* **12**, 4196 (2021). doi: 10.1038/s41467-021-24435-8; pmid: 34234131
38. Y. Cao *et al*., Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* **602**, 657–663 (2022). doi: 10.1038/s41586-021-04385-3; pmid: 35016194
39. A. J. Greaney *et al*., A SARS-CoV-2 variant elicits an antibody response with a shifted immunodominance hierarchy. *PLOS Pathog.* **18**, e1010248 (2022). doi: 10.1371/journal.ppat.1010248; pmid: 35134084
40. P. Wang *et al*., Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **593**, 130–135 (2021). doi: 10.1038/s41586-021-03398-2; pmid: 33684923
41. L. Liu *et al*., Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2. *Nature* **602**, 676–681 (2022). doi: 10.1038/s41586-021-04388-0; pmid: 35016198
42. B. F. Koel *et al*., Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. *Science* **342**, 976–979 (2013). doi: 10.1126/science.1244730; pmid: 24264991
43. N. S. Lewis *et al*., Antigenic and genetic evolution of equine influenza A (H3N8) virus from 1968 to 2007. *J. Virol.* **85**, 12742–12749 (2011). doi: 10.1128/JVI.05319-11; pmid: 21937642
44. N. S. Lewis *et al*., Substitutions near the hemagglutinin receptor-binding site determine the antigenic evolution of influenza A H3N2 viruses in U.S. swine. *J. Virol.* **88**, 4752–4763 (2014). doi: 10.1128/JVI.03805-13; pmid: 24522915
45. N. A. Brasher *et al*., B cell immunodominance in primary hepatitis C virus infection. *J. Hepatol.* **72**, 670–679 (2020). doi: 10.1016/j.jhep.2019.11.011; pmid: 31785346
46. K. Tohma, L. A. Ford-Siltz, J. A. Kendra, G. I. Parra, Dynamic immunodominance hierarchy of neutralizing antibody responses to evolving GII.4 noroviruses. *Cell Rep.* **39**, 110689 (2022). doi: 10.1016/j.celrep.2022.110689; pmid: 35417705
47. L. Popova *et al*., Immunodominance of antigenic site B over site A of hemagglutinin of recent H3N2 influenza viruses. *PLOS ONE* **7**, e41895 (2012). doi: 10.1371/journal.pone.0041895; pmid: 22848649
48. M. Á. Muñoz-Alía, R. A. Nace, L. Zhang, S. J. Russell, Serotypic evolution of measles virus is constrained by multiple co-dominant B cell epitopes on its surface glycoproteins. *Cell Rep. Med.* **2**, 100225 (2021). doi: 10.1016/j.xcrm.2021.100225; pmid: 33948566
49. P. Supasa *et al*., Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera. *Cell* **184**, 2201–2211.e7 (2021). doi: 10.1016/j.cell.2021.02.033; pmid: 33743891
50. D. Zhou *et al*., Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. *Cell* **184**, 2348–2361.e6 (2021). doi: 10.1016/j.cell.2021.02.037; pmid: 33730597
51. C. Liu *et al*., The antibody response to SARS-CoV-2 Beta underscores the antigenic distance to other variants. *Cell Host Microbe* **30**, 53–68.e12 (2022). doi: 10.1016/j.chom.2021.11.013; pmid: 34921776
52. J. Quandt *et al*., Omicron BA.1 breakthrough infection drives cross-variant neutralization and memory B cell formation against conserved epitopes. *Sci. Immunol.* **7**, eabq2427 (2022). doi: 10.1126/sciimmunol.abq2427; pmid: 35653438
53. A. Muik *et al*., Omicron BA.2 breakthrough infection enhances cross-neutralization of BA.2.12.1 and BA.4/BA.5. *Sci. Immunol.* **7**, eade2283 (2022). doi: 10.1126/sciimmunol.ade2283; pmid: 36125366
54. C. I. Kaku *et al*., Recall of preexisting cross-reactive B cell memory after Omicron BA.1 breakthrough infection. *Sci. Immunol.* **7**, eabq3511 (2022). doi: 10.1126/sciimmunol.abq3511; pmid: 35549299
55. Y.-J. Park *et al*., Imprinted antibody responses against SARS-CoV-2 Omicron sublineages. *Science* **378**, 619–627 (2022). doi: 10.1126/science.adc9127; pmid: 36264829
56. S. H. Wilks, acorg/mapping_SARS-CoV-2_antigenic_relationships_and_serological_responses: Mapping SARS-CoV-2 antigenic relationships and serological responses, first release, Version 1.0.0, Zenodo (2023); doi: 10.5281/zenodo.8195839
57. E. J. Anderson *et al*., Safety and Immunogenicity of SARS-CoV-2 mRNA-1273 Vaccine in Older Adults. *N. Engl. J. Med.* **383**, 2427–2438 (2020). doi: 10.1056/NEJMoa2028436; pmid: 32991794
58. L. A. Jackson *et al*., An mRNA Vaccine against SARS-CoV-2 - Preliminary Report. *N. Engl. J. Med.* **383**, 1920–1931 (2020). doi: 10.1056/NEJMoa2022483; pmid: 32663912
59. X. Shen *et al*., SARS-CoV-2 variant B.1.1.7 is susceptible to neutralizing antibodies elicited by ancestral spike vaccines. *Cell Host Microbe* **29**, 529–539.e3 (2021). doi: 10.1016/j.chom.2021.03.002; pmid: 33705729
60. N. A. Doria-Rose *et al*., Booster of mRNA-1273 Strengthens SARS-CoV-2 Omicron Neutralization. medRxiv 2021.12.15.21267805 [Preprint] (2021). https://doi.org/10.1101/2021.12.15.21267805.
61. K. E. Lyke *et al*., Rapid decline in vaccine-boosted neutralizing antibodies against SARS-CoV-2 Omicron variant. *Cell Rep. Med.* **3**, 100679 (2022). doi: 10.1016/j.xcrm.2022.100679; pmid: 35798000
62. W. Li *et al*., Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450–454 (2003). doi: 10.1038/nature02145; pmid: 14647384
63. A. Z. Mykytyn *et al*., Antigenic cartography of SARS-CoV-2 reveals that Omicron BA.1 and BA.2 are antigenically distinct. *Sci. Immunol.* **7**, eabq4450 (2022). doi: 10.1126/sciimmunol.abq4450; pmid: 35737747
64. A. C. Cohen, Progressively Censored Samples in Life Testing. *Technometrics* **5**, 327–339 (1963). doi: 10.1080/00401706.1963.10490102
65. J. J. P. Nauta, Eliminating bias in the estimation of the geometric mean of HI titres. *Biologicals* **34**, 183–186 (2006). doi: 10.1016/j.biologicals.2005.09.001; pmid: 16257535
66. J. Tobin, Estimation of Relationships for Limited Dependent Variables, Estimation of Relationships for Limited Dependent Variables. *Econometrica* **26**, 24–36 (1958). doi: 10.2307/1907382
67. S. H. Wilks, titertools, 2023; https://github.com/shwilks/titertools.
68. J. Kruskal, M. Wish, *Multidimensional Scaling* (Sage, 1978).
69. C. Zhu, R. H. Byrd, P. Lu, J. Nocedal, Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.* **23**, 550–560 (1997). doi: 10.1145/279232.279236
70. S. H. Wilks, Racmacs, version 1.1.35, 14 March 2022, https://acorg.github.io/Racmacs/.
71. S. H. Wilks, "Visualisation and analysis of patterns in serological data using 'antibody landscapes,'" thesis, Univ. of Cambridge (2018).

ACKNOWLEDGMENTS

We thank A. Eaton, E. Domin, S. Hiles, J. Tong, L. Liu, K. Lund, J. Chen, K. Bradley, and W. Feng, Duke University, for immunological analysis. We thank our collaborators on specimen collection and coordination: S. Kaweski, C. Marquez, BC Centre for Disease Control; the Community Surveillance for Infection Team, Epidemiology Task Force, COVID-19 Response (M. Rolfes and A. Dixon), and the Genomics and Discovery Team, Respiratory

Virus Branch DVD, NCIRD from the US Centers for Disease Control and Prevention; the Coronavirus Household Evaluation and Respiratory Testing (C-HEaRT) Cohort Columbia University team (P. Thind, M. Castro, F. Sosa, and L. Wang) and Abt Associates team (J. Duque, Z. Jeddy, K. Altunkaynak, M. Duckworth, and U. Kattel); the CoVPN (COVID-19 Prevention Network) protocol 5001 team; the Natural History of SARS-CoV-2 and Influenza in Peru Study Team; C. Barkauskas, S. Wolery, and I. Naqvi, Duke University Medical Center; the Personalized Virology Initiative at the Icahn School of Medicine at Mount Sinai (D. Altman, M. C. Bermúdez-González, M. Saksena, K. Srivastava, C. Gleason, G. Kleiner, and A. Wajnberg) and the Mount Sinai Pathogen Surveillance Program (A. S. Gonzalez-Reiche, E. M. Sordillo, and A. Paniz-Mondolfi); and J. Beigel and P. C. Roberts from the Division of Microbiology and Infectious Diseases, NIAID, NIH, Bethesda, MD, USA. We are deeply grateful for the generous support of all study participants. The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the [US] Centers for Disease Control and Prevention (CDC). **Funding:** S.H.W., S.T., E.B.L., A.N., R.A.M.F., T.C.J., and D.J.S.: Centers of Excellence for Influenza Research and Response (CEIRR, contract no. 75N93021C00014) as part of the SARS-CoV-2 Assessment of Viral Evolution (SAVE) Program. X.S., D.C.M., C.M., H.T., and L.D.: SAVE Program (contract no. 75N93019C00050 Opt 18A and 18B). B.M., T.C.J., C.D., and V.M.C.: Funded by the German Federal Ministry of Education and Research through project DZIF (8040701710 and 8064701703) and the German Federal Ministry of Health through project SeroVarCoV. V.M.C. and L.M.J.: Funded by the German Federal Ministry of Education and Research through project VARIpath (01KI2021). H.v.B., V.S., and F.K.: NIAID CEIRR contract no. 75N93021C00014, SAVE option 12A, and the NIAID Collaborative Influenza Vaccine Innovation Centers (CIVIC) contract no. 75N93019C00051. E.B.L. is funded by a Gates Cambridge Scholarship. A.N. was supported by the Gates Cambridge Trust. R.P.: Moderna. Also supported by the Office of the Assistant Secretary for Preparedness and Response, Biomedical Advanced Research and Development Authority (contract no. 75A50120C00034) and by NIAID. Y.K.: Research funding support to UW-Madison and University of Tokyo. Center for Research on Influenza Pathogenesis and Transmission (CRIPT) (75N93021C00014) and the Japan Program for Infectious Diseases Research and Infrastructure (JP21wm0125002) from the Japan Agency for Medical Research and Development (AMED). M.S.S.: Funded by the CDC (contract no. 75D30120C08150 with Abt Associates). **Author contributions:** Conceptualization: S.H.W., B.M., X.S., D.C.M., and D.J.S. Methodology: S.H.W. and D.J.S. Software: S.H.W. and T.C.J. Validation: S.H.W., B.M., X.S., S.T., E.B.L., A.N., D.C.M., and D.J.S. Formal analysis: S.H.W., B.M., X.S., S.T., E.B.L., and A.N. Investigation: S.H.W., B.M., X.S., S.T., E.B.L., A.N., D.C.M., and D.J.S. Resources: X.S., M.A.C., J.N.C., M.B.D., T.N.D., C.D., R.A.M.F., P.J.G., V.M.C., L.J., P.J.H., A.J., Y.K., F.K., R.P., V.S., M.S., F.S.D., V.V., H.v.B., R.W., D.C.M., D.J.S., L.M.J., V.M.C., and C.D. Data curation: S.H.W., B.M., and X.S. Writing - Original Draft: S.H.W., B.M., and D.J.S. Writing - Review and Editing: All. Visualization: S.H.W., B.M., S.T., A.N., and E.B.L. Supervision: D.C.M. and D.J.S. **Competing interests:** V.M.C. is named on patents regarding SARS-CoV-2 serological testing and monoclonal antibodies. He is also a part-time employee at Labor Berlin - Charité Vivantes GmbH, a diagnostic laboratory and subsidiary of Charité - Universitätsmedizin Berlin and the Vivantes – Netzwerk für Gesundheit GmbH. F.K. has been consulting for Curevac, Seqirus, and Merck, and is currently consulting for Pfizer, Third Rock Ventures, Avimex, and GSK. He is named on several patents regarding influenza virus and SARS-CoV-2 virus vaccines, influenza virus therapeutics, and SARS-CoV-2 serological tests. Some of these technologies have been licensed to commercial entities, and F.K. is receiving royalties from these entities. F.K. is also an advisory board member of Castlevax, a spin-off company formed by the Icahn School of Medicine at Mount Sinai to develop SARS-CoV-2 vaccines. The Krammer laboratory has received funding for research projects from Pfizer, GSK, and Dynavax, and three of F.K.'s mentees have recently joined Moderna. Y.K. has research funding from FUJIFILM Toyama Chemical Co. (various viruses), Shionogi & Co. (influenza viruses), Daiichi Sankyo Pharmaceutical, Otsuka Pharmaceutical, KM Biologics, Kyoritsu Seiyaku, Fuji Rebio, and Tauns Laboratories. He is a FluGen cofounder and a stockholder of FluGen. **Data and materials availability:** Serum samples from individuals infected with SARS-CoV-2 are available from A. Jassem under a material transfer agreement with the University of British Columbia and the Provincial Health Services Authority. Code and data to reproduce the results and figures in this paper are available from Zenodo (*56*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This research was funded in whole or in part by US govt contract # 75N93021C00014 and the Gates Cambridge Trust, a cOAlition S organizations. The author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adj0070
Figs. S1 to S35
References (*72*–*74*)
Table S1
MDAR Reproducibility Checklist

Submitted 5 June 2023; accepted 23 August 2023
10.1126/science.adj0070

RESEARCH ARTICLE

MATERIALS SCIENCE

Transonic dislocation propagation in diamond

Kento Katagiri[1,2,3,4,5]*, Tatiana Pikuz[6], Lichao Fang[3,4,5], Bruno Albertazzi[7], Shunsuke Egashira[2], Yuichi Inubushi[8,9], Genki Kamimura[1], Ryosuke Kodama[1,2,6], Michel Koenig[1,7], Bernard Kozioziemski[10], Gooru Masaoka[1], Kohei Miyanishi[9], Hirotaka Nakamura[1], Masato Ota[2], Gabriel Rigon[11]†, Youichi Sakawa[2], Takayoshi Sano[2], Frank Schoofs[12], Zoe J. Smith[13], Keiichi Sueda[9], Tadashi Togashi[8,9], Tommaso Vinci[7], Yifan Wang[3,4,5], Makina Yabashi[8,9], Toshinori Yabuuchi[8,9], Leora E. Dresselhaus-Marais[3,4,5], Norimasa Ozaki[1,2]

The motion of line defects (dislocations) has been studied for more than 60 years, but the maximum speed at which they can move is unresolved. Recent models and atomistic simulations predict the existence of a limiting velocity of dislocation motion between the transonic and subsonic ranges at which the self-energy of dislocation diverges, though they do not deny the possibility of the transonic dislocations. We used femtosecond x-ray radiography to track ultrafast dislocation motion in shock-compressed single-crystal diamond. By visualizing stacking faults extending faster than the slowest sound wave speed of diamond, we show the evidence of partial dislocations at their leading edge moving transonically. Understanding the upper limit of dislocation mobility in crystals is essential to accurately model, predict, and control the mechanical properties of materials under extreme conditions.

Motion of dislocations inside a material caused by external stress is related to the material's mechanical properties and its deformation dynamics (*1*, *2*). When a ductile material is stressed, dislocations inside the material move to locally accommodate that force, giving rise to plasticity (*3*). The ductility common to metals is usually absent in brittle materials such as diamond, but even brittle materials can exhibit ductility under some types of extreme conditions like shock-induced high strain-rate deformation (*4*–*6*). Although the basic mechanisms of dislocation-mediated plasticity are sometimes invariant to the rate of strain in the material as it responds to external stresses (*7*), numerous studies have observed cases that show high rate sensitivity (*8*).

At the highest strain rates, dislocations move during deformation at velocities that approach the sound speeds of the material. Dislocation theory predicts that the self-energy and stress of dislocations diverge when a dislocation approaches a limiting (or critical) velocity in a given crystal (*9*–*12*). This implies that dislocations are forbidden to travel at those limiting velocities. Whereas the limiting dislocation velocities in an isotropic crystal coincide with the longitudinal and transverse sound speeds (*13*, *14*), anisotropic single crystals such as diamond have one longitudinal (c_1) and two transverse ($c_1 \geqq c_3$), sound speeds, which are the same order of magnitude but do not always coincide with the limiting velocities (*10*–*12*, *15*–*18*). Critical velocities offset from the sound speeds are attributed to the orientation of a dislocation's motion (*15*). The slowest sound speed, c_3, defines the separation between subsonic and transonic speed regimes.

To date, transonic or supersonic (i.e., faster than c_3 or c_1, respectively) dislocation motion in a real crystal has not been observed experimentally. The only reported experimental evidence of dislocations moving faster than the slowest limiting velocity was in a plasma crystal (*19*). By contrast, numerous theory and molecular dynamics (MD) simulation studies have predicted the existence of transonic or even supersonic dislocation motion, indicating that the limiting velocities should not be the upper limit of dislocation motion (*9*, *20*–*26*). Gumbsch and Gao (*9*) used atomistic simulations to observe ultrafast dislocation motion in tungsten. Their simulations showed that stable motion of dislocations is possible for both transonic and supersonic speeds, but only for dislocations created above the limiting speed, thus avoiding the need to accelerate across the infinite-energy limiting velocity.

Although creating dislocations at such high speeds is relatively straightforward for MD simulations, experiments that measure dislocations have not been able to access those rapidly driven conditions. With >10^7 s^{-1} strain rates, the shock compression technique offers a distinctive system for studying high-velocity dislocations because the energy discontinuity at the shock wavefront can create dislocations that initially move faster than the limiting

[1]Graduate School of Engineering, Osaka University, Suita, 565-0871, Japan. [2]Institute of Laser Engineering, Osaka University, Suita, 565-0871, Japan. [3]Department of Materials Science & Engineering, Stanford University, Stanford, CA 94305, USA. [4]SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. [5]PULSE Institute, Stanford University, Stanford, CA 94305, USA. [6]Institute for Open and Transdisciplinary Research in Initiatives, Osaka University, Suita, 565-0871, Japan. [7]LULI, CNRS, CEA, Ecole Polytechnique, UPMC, Univ Paris 06: Sorbonne Universites, Institut Polytechnique de Paris, Palaiseau, F-91128, France. [8]Japan Synchrotron Radiation Research Institute, Sayo, 679-5198, Japan. [9]RIKEN SPring-8 Center, Sayo, 679-5148, Japan. [10]Lawrence Livermore National Laboratory, Livermore, CA 94550, USA. [11]Department of Physics, Nagoya University, Nagoya, 464-8602, Japan. [12]United Kingdom Atomic Energy Authority, Culham Science Centre, Abingdon OX14 3DB, UK. [13]Department of Applied Physics, Stanford University, Stanford, CA 94305, USA.
*Corresponding author. Email: kentok@stanford.edu
†Present address: Plasma Science and Fusion Center, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

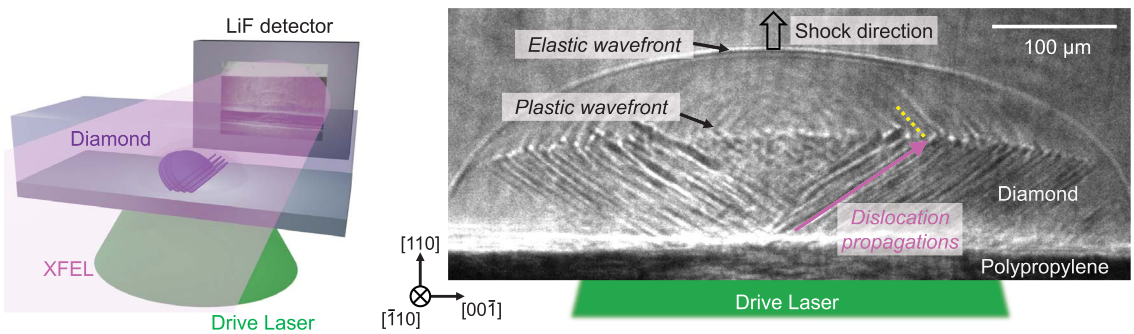


Fig. 1. Femtosecond x-ray radiography on shocked diamond. (**A**) The diamond sample is stressed by shock waves driven by optical laser irradiation. In situ x-ray radiography is measured by using a femtosecond XFEL pulse, irradiating the sample perpendicular to the propagation axis of the shock waves. A lithium fluoride (LiF) crystal is placed at the XFEL downstream to collect the spatial intensity distribution of the transmitted x-rays. (**B**) A representative image that we captured by using the LiF crystal to visualize the elastic-plastic shock wavefronts propagating along the shock direction. Behind the plastic wavefront, stacking faults appear as dark and light bands, indicating the existence of partial dislocations that travel with the plastic shock wavefront (magenta arrow). The yellow dotted line serves to guide the eyes to the radiation.

velocity (*15*, *27*). Here we present experimental results that show shock-induced dislocation motion moving transonically in single-crystal diamond, using femtosecond x-ray radiography. We discuss our interpretation and assignment of the relevant image features and include a discussion of how the dislocation speeds and related plasticity indicate radiation and drag that imply a relationship between the microstructure and bulk elastic-plastic behaviors. Understanding the speed of the fastest dislocation motion is required to accurately predict and control the dislocation dynamics and plasticity of deformed solids, which can be modeled by discrete dislocation dynamics (*28*, *29*). Such ultrafast motion of dislocations strongly affects the mechanical responses of materials in ways that are essential across numerous applications.

Visualization of stacking faults

Our experiments were conducted using the X-ray Free Electron Laser (XFEL) at SPring-8 Angstrom Compact Free Electron Laser (SACLA). At SACLA's Experimental Hutch 5, a high-intensity nanosecond-pulsed optical laser is aligned to the same spatial position as the femtosecond-pulsed XFEL, and its timing is synchronized to perform in situ x-ray measurements of laser-shocked materials (*30*). We used type IIa single crystals of diamond synthesized by chemical vapor deposition. Our diamonds were cut along two different orientations to enable our shock to propagate along the [100] and [110] directions, to examine the orientational difference in the plasticity mediated by shock-induced defects (see fig. S9 for [111] shock data). We estimate the peak shock stresses applied to the diamond samples to be 184 ± 16 GPa and 92 ± 15 GPa for [100] and [110] shock directions, respectively (*18*). These shock stresses are high enough to make diamond yield (*5*), generating strong shear stresses that initiate failure mechanisms. When diamond yields during shock compression, the shock wave's front splits into two: a leading elastic front that is followed by the plastic front in which dislocation propagation occurs.

Although several x-ray imaging techniques are used for in situ measurements (*31*–*34*), x-ray radiography is one of the original ones used to nondestructively visualize structures in bulk materials (*35*). X-ray radiography spatially maps the x-ray intensity transmitted through a material of interest, giving image contrast that is usually based on the absorption by the sample. The contrast that we observe in this work is mainly due to the scattering of the x-rays rather than absorption; despite the similarity in contrast mechanism, our radiography setup is distinct from the phase contrast imaging setups used in other XFEL shock experiments (*32*). To perform the in situ x-ray radiography, we used an unfocused and single-pulsed XFEL beam with a photon energy of 7.0 keV and pulse duration of ~8 fs to illuminate the diamond along the axis perpendicular to the shock propagation direction (Fig. 1). We placed a lithium fluoride (LiF) crystal downstream of the sample along the XFEL beam as a vacuum-compatible imaging detector that offers a $>10^6$ dynamic range and ~1-μm spatial resolution over a wide field of view (*36*). The distance between the diamond sample and the front surface of the LiF crystal was 112 mm.

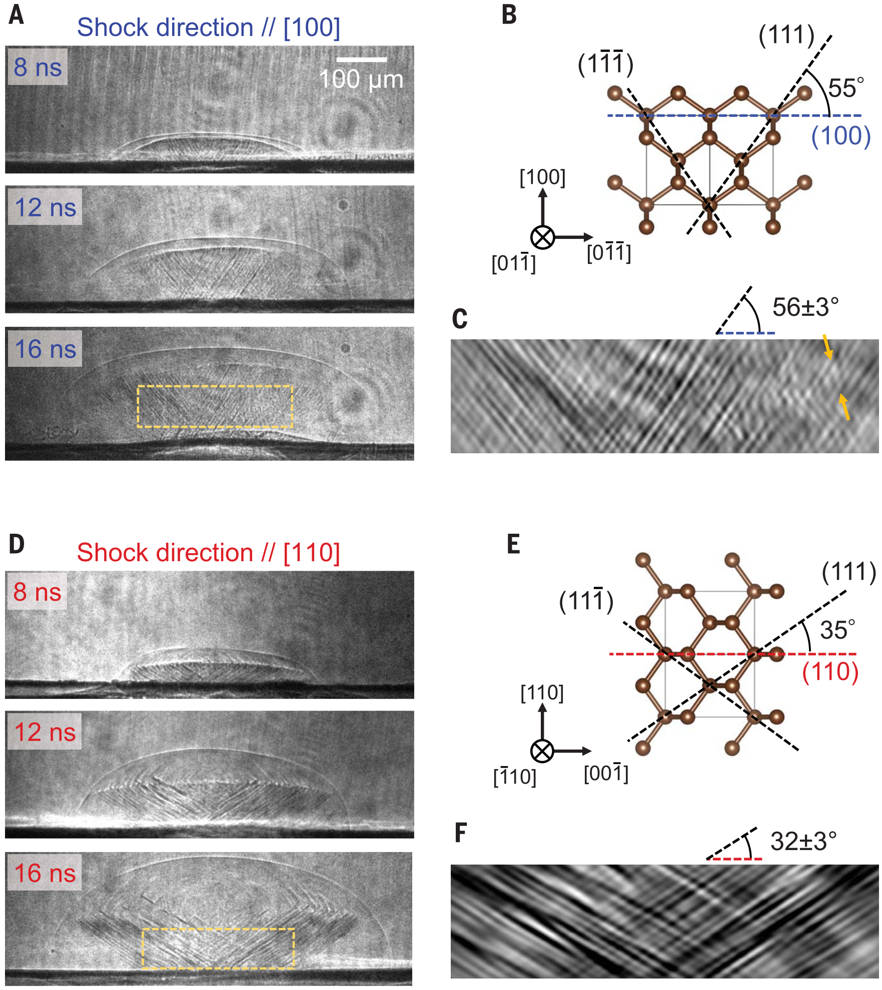


Fig. 2. Visualization of stacking faults. (**A**) X-ray radiography images of the diamond shocked along the [100] direction. The numbers on each image indicate the XFEL probe timing relative to the drive laser irradiation. The vertical lines that appear across the entire field of view are the texture on the unpolished lateral surfaces of the diamond. The two large circles with concentric scattering patterns are caused by scattering from the debris on a beryllium window upstream of the XFEL. (**B**) Undistorted diamond lattice structures (*37*) with the corresponding plane orientations. (**C**) Fast Fourier transform (FFT)–filtered image showing the stacking faults observed in the 16-ns shot. The corresponding area is indicated as the yellow dashed rectangle in (A). The orange arrows serve to guide the eyes to one of the bands formed nonparallel to the diamond {111} planes (see text). (**D**) X-ray radiography images when diamonds are shocked along [110] direction. The 12-ns image is identical to the one shown in Fig. 1B. (**E**) Undistorted lattice structure showing the plane orientations. (**F**) FFT-filtered image of the [110] shock at 16 ns. The errors shown on the angles noted in (C) and (F) represent 1σ of multiple measurements (*18*).

Our x-ray radiography for a diamond shocked along the [110] direction captures two shock wavefronts, corresponding to a preceding elastic wave and a following plastic wave, traversing the diamond sample from the bottom to the top of the image (Fig. 1). Behind the plastic wavefront, light and dark bands appear diagonal to the shock direction. In all cases, the two characteristic vectors of the shock wave and the XFEL imaging beam's direction indicate the angles of the linear banded features appear parallel to the {111} planes of the crystal. When the {111} planes are set not along the XFEL's probing direction (i.e., the sample was placed 90° rotated about the shock propagation axis), these banded features appear much fainter (fig. S10). This indicates that the image contrast of the banded

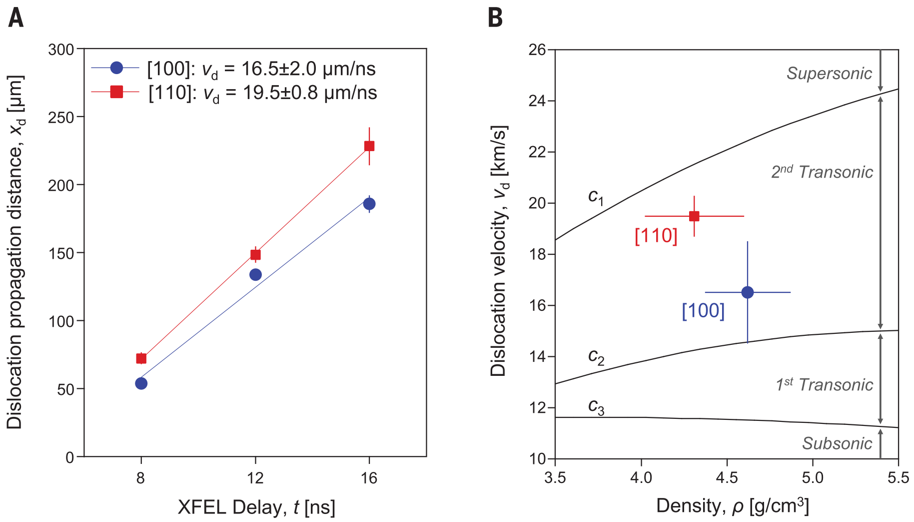


Fig. 3. Transonic dislocation motion in shocked diamonds. (**A**) Measured dislocation propagation distances in diamond (x_d) as a function of the XFEL delay (t). The x_d = 0 μm denotes the interface between the ablator and diamond, whereas t = 0 ns denotes the time that the laser hits the ablator. The blue circles and red squares represent our experimental results for the [100] and [110] shock directions, respectively. The velocities of the dislocation propagation (v_d) noted are obtained from the slope of the linear fits (solid lines). The error bars on each plot represent 1σ of multiple measurements. The shot-by-shot fluctuation of drive laser intensity is not included in the error bars shown on x_d. The errors on v_d are evaluated on the basis of the 1σ of the fitting and the errors on each plot. (**B**) v_d versus the diamond's material density (ρ) at the shocked states. Black curves are the calculated c_1, c_2, and c_3 of diamond propagating along [110] direction (*18*). The error bars on ρ are propagated from the errors on the plastic shock wave velocities (*18*).

features in our x-ray radiography is strongest when the x-ray beam direction integrates along a plane perpendicular to the two-dimensional (2D) features formed in the diamond.

We used Fourier filtering methods to emphasize the linear image features (Fig. 2, C and F). These banded features in the [110] shock images appear thicker and darker than those seen in the [100] shock images. As described in the next section, the dislocation propagation for the [110] shock is faster than that for the [100] shock, which potentially affects the thickness and darkness of the stacking faults as a faster dislocation propagation results in faster macroscopic plastic flow (*38*), causing more damage to the material. We measured the angles between the linear features and the diamond surface using Hough transforms on the Fourier-filtered images (*18*). The observed angles of the 16-ns images are 56° ± 3° and 32° ± 3° for [100] and [110] shocks, respectively. Whereas the value for [100] agrees with the angles between the diamond surface and the {111} plane of the undistorted diamond lattice, which is 55° (Fig. 2B), the observed angle for [110] shows a relatively larger deviation from the angle of the undistorted diamond lattice, which is 35° (Fig. 2E). Although the deviation is within the estimated error, we believe this deviation suggests that the diamond shocked along [110], experiencing a plastic deformation with some distortion caused by the uniaxial shock compression. We observed bands along non-{111} directions only in the [100] shock image collected at 16-ns delay (Fig. 2C, orange arrows). The angular offset may indicate shock-induced twinning or slip along another crystallographic plane, though we cannot exclude the possibility of the diamond crystal having an undesirable formation of a rotated grain during the synthesis.

We interpret these deformation-induced bands as accumulated stacking faults (*39*) that form along the {111} planes. The {111} plane in diamond is known to be the dominant slip plane for dislocations. Previous shock-loading experiments in other brittle materials have reported the occurrence of shock-induced amorphous bands formed diagonally to the shock wavefront (*40*–*44*), but the estimated temperature in the shocked diamond in our work is less than 20% of its melting temperature (*18*), which we interpret as too low for amorphization. The possibility of phase transformation is also negligible because the applied stress in our experiments is far below the phase transition point (*45*, *46*).

Transonic dislocation propagation

Elastic theory predicts that transonic dislocations can be generated at discontinuities in energy or displacement in a material, as would be present at a shock-wave front (*15*, *27*). These conditions can cause a full dislocation to split into partial dislocations, leaving a stacking fault between them (*47*, *48*). As the edges of a stacking fault are defined by partial dislocations (*49*, *50*), the leading edge of the stacking-fault extension that we observe gives the dislocation velocity in our experiment. In our images, the stacking faults formed in the diamonds appear as discontinuous lines, indicating that the dislocations travel with the plastic wavefront, originating from the laser-irradiated side of the diamond surface. A closer look at the [110] shock images (Fig. 1B) shows the absence of these bands at the center of the plastic wavefront. The reason for this observation is that most of the dislocations traveling with the plastic wavefront are propagated from the limited area of the diamond surface that is shocked by the drive laser (φ ~ 260 μm). This indicates that the number of dislocations freshly created at the propagating plastic shock front is much less than those propagating from the diamond surface.

We plot the length of the stacking faults that we observe in our images (x_d) as a function of time delay after the shock initiation (t) (Fig. 3A). Then we quantify the velocity of the partial dislocations (v_d) leading that stacking-fault extension by fitting the points to a line with $v_d = dx_d/dt$ from the measured x_d-t relationship. We show the obtained dislocation velocities (Fig. 3B) along with the sound-wave velocities (c_1, c_2, and c_3) of diamond at various densities (*18*) calculated from the pressure-dependent elastic constants at 1100 K from (*51*). The dislocation velocities that we observe for the [100] and [110] shock directions both lie in the transonic regime (i.e., between c_1 and c_3). We find that the velocity of dislocations in the [110] shocked crystals is faster than those of the [100] crystal. Although the plastic wave velocities differ only slightly, the angle between the shock direction and dislocation propagation direction is larger for the [110] shock than the [100] shock, making the dislocation propagation distance longer (and thus faster) for the [110] to catch up with the plastic shock wavefront.

The existence of supersonic dislocation has been observed in several MD simulations though some of the early theoretical studies predicted it to be impossible (*27*, *52*). Because the shock-induced dislocation motion that we observe in diamond travel with the plastic shock wavefront, applying a higher shock stress to drive a faster plastic shock wave can potentially drive even faster dislocation motion. This will allow one to experimentally investigate the possible existence of supersonic dislocations in a real crystal.

Radiation from transonic dislocations

A close look at the elastically compressed volume immediately preceding the plastic wavefront reveals the presence of multiple "elastic pulses" (Fig. 1B, yellow dotted line). These pulses are referred to as radiation (or phonon radiation) in dislocation theory (*27*). Although the emission of radiation from transonically

moving dislocations has been predicted by dislocation theory and observed in MD simulations (*9*, *23*, *24*), a dearth of validation methodology has precluded experimental observation of the radiation.

Although we observe the radiation feature in our images of the [110] shock (Fig. 2D), our images for the [100] (Fig. 2A) and [111] shock directions (fig. S9) show less clear signs of radiation. The different amounts of radiation that we observe for shocks along different directions with different dislocation velocities should be due to the strong velocity dependence on the energy dissipation in radiation, which has also been predicted by the theory and simulations (*9*, *53*). Because the radiation from the transonically propagating dislocation cores exerts a negative driving force that adds drag to the dislocation motion, large external stress is required to drive the stable motion of transonic dislocations without deceleration (*9*). The use of strong shock waves can provide stress that is high enough to overcome the drag by radiation to achieve stable transonic motion of dislocations.

Although the existence of the radiation from transonically propagating dislocation cores and its role in the energy dissipation have been extensively studied, its effect on the shock-induced deformation dynamics remains largely unknown. A few studies have used MD simulations of shocked single-crystal aluminum (*54*, *55*) to predict shock-induced elastic-plastic deformation mechanisms wherein the preceding elastic volume was disturbed by the plasticity that followed. Their observations were explained by the elastic pulses emitted from the dislocations at the plastic wavefront, consistent with the configuration that we observe in this diamond study. A recent billion-atom MD simulation of single-crystal diamond shocked along the [110] axis revealed complex defect structures at a peak stress that was slightly higher than 400 GPa (*56*). Such dislocation-driven elements of shock deformation dynamics might explain the complex elastic-plastic correlations in shocked diamond (*57*) and the anomalous elastic behavior in shocked silicon (*58*) that were observed in previous experiments by using a conventional velocimetry.

In conclusion, our in situ x-ray radiography using one of the world's brightest x-ray lasers shows experimental evidence of transonic dislocation motion in diamond, leading to the formation of numerous stacking faults that span the plastically deformed volume. The microscopic dislocation motions that emit radiation might affect the macroscopic elastic-plastic deformation dynamics. Our experimental results showing transonic dislocation motion open key new opportunities to refine models for insights into the ultrafast deformation behaviors at these extreme conditions. The newly refined models at the highest strain rates will have a marked influence on numerous fields, including ultrafast fracture in structural materials (*59*–*61*), prediction and analysis of earthquake ruptures (*62*–*65*), precise manufacturing processes (*66*), and functionalities in electrochemical applications (*67*).

REFERENCES AND NOTES

1. W. G. Johnston, J. J. Gilman, *J. Appl. Phys.* **31**, 632–643 (1960).
2. M. A. Meyers, H. Jarmakani, E. M. Bringa, B. A. Remington, *Dislocations in Solids* (Elsevier, 2009).
3. R. Madec, B. Devincre, L. Kubin, T. Hoc, D. Rodney, *Science* **301**, 1879–1882 (2003).
4. G. I. Kanel, V. E. Fortov, S. V. Razorenov, *Shock-Wave Phenomena and the Properties of Condensed Matter* (Springer-Verlag, 2004).
5. R. S. McWilliams *et al*., *Phys. Rev. B Condens. Matter Mater. Phys.* **81**, 014111 (2010).
6. M. J. MacDonald *et al*., *Appl. Phys. Lett.* **116**, 234104 (2020).
7. L. A. Zepeda-Ruiz, A. Stukowski, T. Oppelstrup, V. V. Bulatov, *Nature* **550**, 492–495 (2017).
8. H. Fan, Q. Wang, J. A. El-Awady, D. Raabe, M. Zaiser, *Nat. Commun.* **12**, 1845 (2021).
9. P. Gumbsch, H. Gao, *Science* **283**, 965–968 (1999).
10. L. J. Teutonico, *Phys. Rev.* **124**, 1039–1045 (1961).
11. L. J. Teutonico, *Phys. Rev.* **127**, 413–418 (1962).
12. D. N. Blaschke, J. Chen, S. Fensin, B. A. Szajewski, *Philos. Mag.* **101**, 997–1018 (2021).
13. J. Weertman, J. R. Weertman, in Moving Dislocations, F. R. N. Nabarro, ed., vol. 3 of Dislocations in Solids, pp. 1–59, (Amsterdam: North Holland Pub. Co., 1980).
14. J. Eshelby, *Proc. Phys. Soc. A* **62**, 307–314 (1949).
15. A. N. Stroh, *J. Math. Phys.* **41**, 77–103 (1962).
16. D. N. Blaschke, *J. Phys. Condens. Matter* **33**, 503005 (2021).
17. D. N. Blaschke *et al*., On radiation-free velocities of edge dislocations in arbitrary slip systems of anisotropic crystals, arXiv:2305.06980v1 (2023).
18. Materials and methods are available as supplementary materials.
19. V. Nosenko, S. Zhdanov, G. Morfill, *Phys. Rev. Lett.* **99**, 025002 (2007).
20. H. Tsuzuki, P. S. Branicio, J. P. Rino, *Appl. Phys. Lett.* **92**, 191909 (2008).
21. Z. Jin, H. Gao, P. Gumbsch, *Phys. Rev. B Condens. Matter Mater. Phys.* **77**, 094303 (2008).
22. S. Peng, Y. Wei, Z. Jin, W. Yang, *Phys. Rev. Lett.* **122**, 045501 (2019).
23. P. Gumbsch, H. Gao, *J. Comput. Aided Mater. Des.* **6**, 137–144 (1999).
24. J. Weertman, *J. Appl. Phys.* **38**, 5293–5301 (1967).
25. T. Duong, M. J. Demkowicz, *J. Mech. Phys. Solids* **180**, 105422 (2023).
26. B. Gurrutxaga-Lerma, J. Verschueren, A. P. Sutton, D. Dini, *Int. Mater. Rev.* **66**, 215–255 (2020).
27. P. M. Anderson *et al*., *Theory of Dislocations* (Cambridge Univ. Press, 2017).
28. B. Gurrutxaga-Lerma, D. S. Balint, D. Dini, D. E. Eakins, A. P. Sutton, *Proc. R. Soc. A* **469**, 20130141 (2013).
29. N. Bertin, R. B. Sills, W. Cai, *Annu. Rev. Mater. Res.* **50**, 437–464 (2020).
30. Y. Inubushi *et al*., *Appl. Sci.* **10**, 2224 (2020).
31. A. Schropp *et al*., *Sci. Rep.* **5**, 11089 (2015).
32. F. Seiboth *et al*., *Appl. Phys. Lett.* **112**, 221907 (2018).
33. S. B. Brown *et al*., *Sci. Adv.* **5**, eaau8044 (2019).
34. L. E. Dresselhaus-Marais *et al*., *Sci. Adv.* **7**, eabe8311 (2021).
35. M. Moore, *J. Phys. Condens. Matter* **21**, 364217 (2009).
36. T. Pikuz *et al*., *Matter Radiat. Extremes* **3**, 197–206 (2018).
37. K. Momma, F. Izumi, *J. Appl. Cryst.* **44**, 1272–1276 (2011).
38. E. Orowan, *Proc. Phys. Soc.* **52**, 8–22 (1940).
39. B. L. Holian, P. S. Lomdahl, *Science* **280**, 2085–2088 (1998).
40. M. Chen, J. W. McCauley, K. J. Hemker, *Science* **299**, 1563–1566 (2003).
41. S. Zhao *et al*., *Acta Mater.* **103**, 519–533 (2016).
42. S. Zhao *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12088–12093 (2016).
43. S. Zhao *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **114**, 9791–9796 (2017).
44. S. Zhao *et al*., *Acta Mater.* **158**, 206–213 (2018).
45. M. D. Knudson, M. P. Desjarlais, D. H. Dolan, *Science* **322**, 1822–1825 (2008).
46. A. Lazicki *et al*., *Nature* **589**, 532–535 (2021).
47. S. Masuya, K. Hanada, T. Oshima, H. Sumiya, M. Kasu, *Diam. Relat. Mater.* **75**, 155–160 (2017).
48. J. A. Y. Vandersall, B. D. Wirth, *Philos. Mag.* **84**, 3755–3769 (2004).
49. W. Cai, W. D. Nix, *Imperfections in Crystalline Solids* (Cambridge Univ. Press, 2016).
50. D. Hull, D. J. Bacon, *Introduction to Dislocations* (Elsevier, 2011).
51. M. Núñez Valdez, K. Umemoto, R. M. Wentzcovitch, *Appl. Phys. Lett.* **101**, 171902 (2012).
52. J. Friedel, *Dislocations* (Pergamon, Oxford, 1964).
53. H. Gao, Y. Huang, P. Gumbsch, A. J. Rosakis, *J. Mech. Phys. Solids* **47**, 1941–1961 (1999).
54. V. V. Zhakhovsky, M. M. Budzevich, N. A. Inogamov, I. I. Oleynik, C. T. White, *Phys. Rev. Lett.* **107**, 135502 (2011).
55. V. V. Zhakhovsky, N. A. Inogamov, B. J. Demaske, I. I. Oleynik, C. T. White, *J. Phys. Conf. Ser.* **500**, 172007 (2014).
56. K. Nguyen-Cong *et al*., *In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, 1–12 (2021).
57. K. Katagiri *et al*., *Phys. Rev. Lett.* **125**, 185701 (2020).
58. A. Loveridge-Smith *et al*., *Phys. Rev. Lett.* **86**, 2349–2352 (2001).
59. A. J. Rosakis, O. Samudrala, D. Coker, *Science* **284**, 1337–1340 (1999).
60. M. J. Buehler, F. F. Abraham, H. Gao, *Nature* **426**, 141–146 (2003).
61. M. J. Buehler, H. Gao, *Nature* **439**, 307–310 (2006).
62. M. Bouchon, M. Vallée, *Science* **301**, 824–826 (2003).
63. M. Bouchon, H. Karabulut, *Science* **320**, 1323–1325 (2008).
64. E. M. Dunham, P. Favreau, J. M. Carlson, *Science* **299**, 1557–1559 (2003).
65. S. Das, *Science* **317**, 905–906 (2007).
66. Y. Liao *et al*., *J. Appl. Phys.* **110**, 023518 (2011).
67. K. K. Adepalli, J. Yang, J. Maier, H. L. Tuller, B. Yildiz, *Adv. Funct. Mater.* **27**, 1700243 (2017).

ACKNOWLEDGMENTS

We thank H. Sumiya, Y. Seto, T. Okuchi, and T. Duong for discussions and technical staff of SACLA for their support during the experiments. The experiments were performed at BL3 of SACLA with the approval of the Japan Synchrotron Radiation Research Institute (proposal nos. 2021A8004 and 2021B8067). The high-power nanosecond laser at SACLA was installed by the Institute of Laser Engineering of Osaka University with the corporation of Hamamatsu Photonics. The installation of Diffractive Optical Elements (DOE) to improve the smoothness of the drive laser pattern was supported by the SACLA Basic Development Program. **Funding:** MEXT Quantum Leap Flagship Program (MEXT Q-LEAP) grant no. JPMXS0118067246 Japan Society for the Promotion of Science (JSPS) KAKENHI grant nos. 21K03499, 21J10604, and 20H00139 JSPS Core-to-Core Program B: Asia-Africa Science Platforms grant no. JPJSCCB20190003 Genesis Research Institute, Inc. (Konpon-ken, Toyota) K.K. acknowledges support from Yamada Science Foundation. B.K. acknowledges support from US Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. K.K. and L.E.D.-M. acknowledge support from Air Force Office of Scientific Research "FA9550-23-1-0347," managed by C. Li and D. Barbee. Y.W. was supported by the Stanford Energy Postdoctoral Fellowship through contributions from the Precourt Institute for Energy, Bits & Watts Initiative, StorageX Initiative, and TomKat Center for Sustainable Energy. **Author contributions:** Conceptualization: K.K., N.O.; Methodology: K.K., T.P., L.E.D.-M.; Resources: F.S.; Investigation: K.K., T.P., B.A., S.E., Y.I., G.K., M.K., G.M., K.M., H.N., M.O., G.R., Y.S., T.S., Z.J.S., K.S., T.T., T.V., M.Y., T.Y., L.E.D.-M., N.O.; Visualization: K.K., L.F., Z.J.S.; Formal Analysis: K.K., T.P., L.F., B.K., Y.W.; Funding acquisition: K.K., L.E.D.-M., N.O., Y.W.; Project administration: N.O.; Supervision: R.K., L.E.D.-M. Writing – original draft: K.K., L.E.D.-M.; Writing – review and editing: All authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. Raw images are available on the Open Science Framework repository (https://osf.io/cb4e2/?view_only=6db54c9342a54db8b8681f9898bea7ea).

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh5563
Materials and Methods
Supplementary Text
Figs. S1 to S12
Table S1
References (*68*–*78*)

Submitted 11 March 2023; accepted 16 August 2023
10.1126/science.adh5563

HUMAN EVOLUTION

Independent age estimates resolve the controversy of ancient human footprints at White Sands

Jeffrey S. Pigati[1]*†, Kathleen B. Springer[1]*†, Jeffrey S. Honke[1], David Wahl[2,3], Marie R. Champagne[2], Susan R. H. Zimmerman[4], Harrison J. Gray[1], Vincent L. Santucci[5], Daniel Odess[6]‡, David Bustos[7], Matthew R. Bennett[8]

Human footprints at White Sands National Park, New Mexico, USA, reportedly date to between ~23,000 and 21,000 years ago according to radiocarbon dating of seeds from the aquatic plant *Ruppia cirrhosa*. These ages remain controversial because of potential old carbon reservoir effects that could compromise their accuracy. We present new calibrated ^{14}C ages of terrestrial pollen collected from the same stratigraphic horizons as those of the *Ruppia* seeds, along with optically stimulated luminescence ages of sediments from within the human footprint–bearing sequence, to evaluate the veracity of the seed ages. The results show that the chronologic framework originally established for the White Sands footprints is robust and reaffirm that humans were present in North America during the Last Glacial Maximum.

Ancient footprints in White Sands National Park, New Mexico, USA (WHSA), appear to place humans in North America during the Last Glacial Maximum (*1*). The footprints and associated tracks of Pleistocene megafauna at WHSA Locality 2 are imprinted on multiple stratigraphic horizons composed of fine-grained, gypsum-rich alluvium intercalated with clay and silt that was deposited in a mosaic of wet and dry environments along the eastern margin of Paleolake Otero in the Tularosa Basin (Fig. 1). Seeds from the aquatic plant *Ruppia cirrhosa* are found in situ within thin clay laminae, either interbedded with the footprint horizons or embedded in the footprints themselves. A total of 11 aliquots of *Ruppia* seeds yielded calibrated ^{14}C ages that maintain stratigraphic order and range from ~23 to 21 ka (ka, thousand years before present; 0 years = 1950 CE) (*1*), constraining the human footprints to within the temporal bounds of the Last Glacial Maximum (~26.5 to 19 or 20 ka) (*2*).

The findings at WHSA Locality 2 push back the peopling of the Americas by thousands of years and imply that early inhabitants and megafauna coexisted for several millennia before the terminal Pleistocene extinction event (*3*). This has important consequences for archaeology and allied fields, yet there has been widespread debate as to the accuracy of the *Ruppia* ages and therefore the antiquity of the footprints. This debate is organized around two lines of reasoning. The first is that *R. cirrhosa* is an aquatic plant, so there is the potential for uptake of old carbon from groundwater, which would cause the seed ages to be too old, a phenomenon known as the "hard-water effect" (*4*, *5*). The second is that *Ruppia* seeds are physically robust and can be preserved in sediments over geologic timescales, so it is possible that the seeds were exhumed from older deposits and reworked into the sediments at WHSA Locality 2. If so, the seeds would provide only maximum-limiting ages for the human footprints (*6*, *7*).

To address the controversy regarding the ages derived from *Ruppia* seeds, we obtained radiocarbon ages of terrestrial pollen recovered from the same stratigraphic intervals as those of the seeds, as well as optically stimulated luminescence (OSL) ages of quartz grains from within the footprint horizon interval, to evaluate the chronology of WHSA Locality 2. The dating techniques, the sample types, and the accelerator mass spectrometry facility involved in this study are independent of those used by Bennett *et al.* (*1*).

Terrestrial pollen is produced by cone-bearing and fruit-bearing plants living on land and fixing atmospheric carbon and therefore is not subject to hard-water effects. However, pollen grains are extremely lightweight and are generally between 10 and 150 μm in diameter (*8*), meaning that tens or even hundreds of thousands of grains are required to achieve the mass necessary for a single radiocarbon measurement. Even though it is an exceptionally laborious process, researchers have attempted to date pollen for decades (*9*–*17*), but the results have been mixed largely because of the difficulty of isolating enough pollen grains from other organic material to produce reliable ages. The recent adoption of flow cytometry to separate pollen grains from sediment and other organic materials has overcome this



Fig. 1. Ancient human footprints found at WHSA Locality 2. Photograph showing multiple in situ footprints in track horizons TH1 (white arrows), TH2 (yellow arrows), TH3 (red arrows), and TH4 (black arrows). Their stratigraphic positions and associated chronologic data are provided in Fig. 2.

[1]US Geological Survey, Denver Federal Center, Denver, CO 80225, USA. [2]US Geological Survey, Menlo Park, CA 94025, USA. [3]Department of Geography, University of California, Berkeley, Berkeley, CA 94720, USA. [4]Center for Accelerator Mass Spectrometry, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA. [5]National Park Service, Geologic Resources Division, Washington, DC 20240, USA. [6]National Park Service, Cultural Resources Directorate, Washington, DC 20240, USA. [7]National Park Service, White Sands National Park, Holloman Air Force Base, NM 88330, USA. [8]Department of Life and Environmental Sciences, Bournemouth University, Poole BH12 5BB, UK.
***Corresponding author. Email: jpigati@usgs.gov (J.S.P.); kspringer@usgs.gov (K.B.S.)**
†These authors contributed equally to this work.
‡Present address: University of Alaska Museum of the North, Fairbanks, AK 99775, USA.

problem, and pollen samples often approach purity after processing (*18*–*21*).

In this study, we collected large (>1 kg) samples of bulk sediment from the exact same stratigraphic levels as those of the original samples of *Ruppia* seeds from WHSA Locality 2, which yielded ages that range from 22.87 ± 0.30 to 21.13 ± 0.25 ka (Fig. 2) (*1*). The samples were subjected to rigorous chemical pretreatment and microsieving to remove as much of the host material as possible, and the pollen grains were separated from the remaining residue by means of flow cytometry (*22*). All of the WHSA Locality 2 samples contain pollen assemblages that are indicative of a climate that was much cooler and wetter than today (*22*). Arboreal taxa include abundant *Pinus* (pine) and some *Picea* (spruce), *Abies* (fir), and *Pseudotsuga* (Douglas Fir), which indicate altitudinal lowering of nearby conifer forests compared with present day and are consistent with previous findings for the Last Glacial Maximum (*23*). Nonarboreal taxa are dominated by *Artemisia*, reflecting a sagebrush steppe that is not found in the region today (*22*). For dating, we targeted conifer pollen because it is relatively large (>70 μm) and has a relatively thick exine, both of which result in increased carbon content. We measured the ^{14}C content of three different samples from WHSA Locality 2 that each contained ~75,000 pollen grains, and the resulting calibrated ages range from 23.4 ± 2.5 ka to 22.6 ± 2.3 ka (Fig. 2) (*22*).

The uncertainties associated with the pollen ages are relatively large because they required blank corrections that are approximately an order of magnitude larger than what is typically used in radiocarbon dating. Blank corrections are performed on all radiocarbon samples to account for small amounts of contaminant ^{14}C (usually modern) that is introduced to samples during the various stages of processing, including collection, handling, chemical pretreatment, CO_2 extraction, graphitization, and the isotopic measurement (*24*–*26*). To isolate the pollen grains, the samples analyzed in this study required extensive pretreatment procedures, so in order to properly correct the resulting ages, we also extracted and dated pollen from sediments that are known to be beyond the limit of radiocarbon dating for use as a procedural blank (*22*). Assuming that the same amount of contamination was introduced to all of the samples analyzed in this study, which is equivalent to an increase of ~2200 ^{14}C years or 2500 calendar years for the WHSA Locality 2 sequence (*22*), the resulting ages should be viewed as being accurate, albeit with reduced precision. Overall, the calibrated ^{14}C ages of the pollen samples from WHSA Locality 2 are statistically indistinguishable from the original seed ages (Fig. 3), which supports the original chronologic framework developed by Bennett *et al.* (*1*). Last, even had we not undertaken this pollen-specific approach for the blank correction, the calibrated pollen ages would still have fallen within the Last Glacial Maximum (*22*).

In addition to radiocarbon methods, which are based on radioactive decay, we used OSL dating, which is based on the accumulation of energy stored as trapped charge within the crystal structure of a mineral and is a function of background ionizing radiation (*27*). This trapped charge is released and the geochronometer is reset when a mineral, typically quartz, is exposed to sunlight or heat. For the sediments at WHSA Locality 2, an OSL age is equivalent to the amount of time elapsed since the quartz was deposited and sealed from sunlight. Although the sediments at WHSA Locality 2 are dominated by gypsum and contain only a small amount of quartz, we were able to obtain three OSL age minima from a single stratigraphic level located just below our lowest radiocarbon age but still well within the human footprint horizons. Collectively, they show the minimum OSL age of the sampled horizon is >21.5 ± 1.9 ka (*22*). As with the pollen results, the OSL ages are statistically indistinguishable from the calibrated ^{14}C ages of the lowest *Ruppia* seeds collected for dating from the stratigraphic sequence (Fig. 2).

Both the calibrated ^{14}C ages of pollen and the OSL ages support the original ages obtained from the *Ruppia* seeds at WHSA Locality 2, but the accuracy of the age estimates from each method must be evaluated individually before they can be accepted as reliable. The most

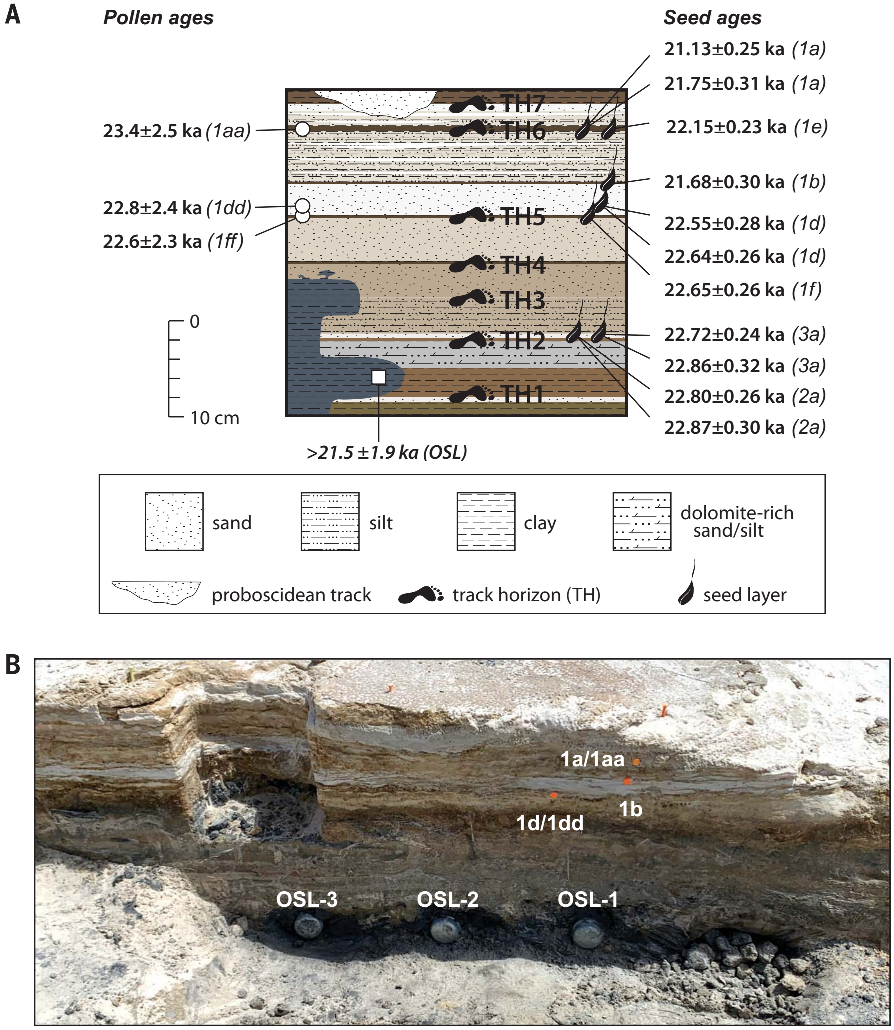


Fig. 2. Geologic context of chronologic samples. (**A**) Composite stratigraphy, calibrated ^{14}C ages, and sample numbers (in italics) of pollen and *Ruppia* seeds, as well as the minimum OSL age for sediments exposed in the trench at WHSA Locality 2. The calibrated ^{14}C ages were used to construct the age models shown in Fig. 3. (**B**) Photograph looking south of the sediments exposed in the trench at WHSA Locality 2 along with locations of radiocarbon samples for *Ruppia* seeds (1a, 1b, and 1d) and pollen (1aa and 1dd) as well as the three OSL samples (OSL-1, -2, and -3) (*22*). Pollen sample 1ff is located near the base of the trench to the east (fig. S1). The large block cut out of the exposed face on the left side of the photograph was sampled previously for U-Th series disequilibrium dating (*1*).

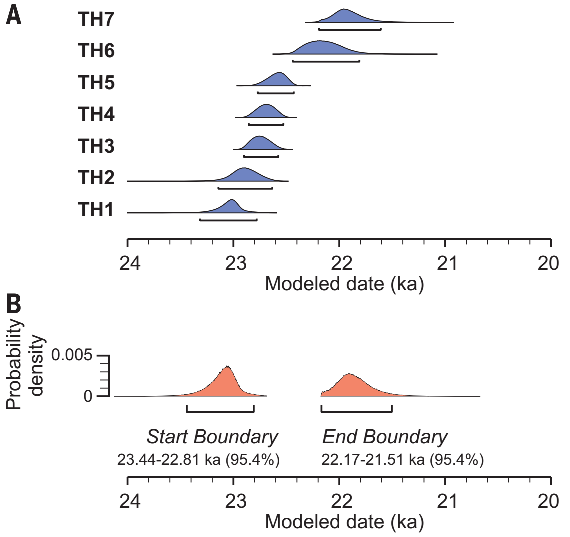


Fig. 3. Age dating and modeling results. Modeled ages for (**A**) track horizons TH1 through TH7 and (**B**) the start and end boundaries for the entire footprint horizon interval based on the calibrated ^{14}C ages of pollen and *Ruppia* seeds. Both figures were generated using the Sequence function in OxCal v.4.4.4 (*32*) and the IntCal20 calibration curve (*23*), which we accessed online 30 May 2023. Samples identified as outliers at the 95% confidence level were excluded from the analyses (*22*).

common pitfall when dating pollen is that reworking of pollen from older sediments could potentially cause the ages to be too old (*19*, *28*). To evaluate this possibility, we dated multiple aliquots of pollen extracted from sediments collected at the surface of an active playa in White Sands National Park. The pollen assemblage in the modern playa sediments is consistent with the warm, dry conditions that prevail in the Tularosa Basin today and is composed primarily of nonarboreal taxa, including Amaranthaceae, Asteraceae, and Poaceae (all flowering plants) (*22*). The ^{14}C content of the playa conifer pollen (*22*) is indistinguishable from current atmospheric ^{14}C levels (*29*), which shows that reworking of old pollen is inconsequential in this part of the Tularosa Basin today. It follows that this also holds true for the late Pleistocene, especially considering that much of the basin was covered by Paleolake Otero at that time.

Similarly, processes that take place after sediments are deposited can potentially influence OSL ages by affecting the background radiation field, causing them to be either too old or too young. Such phenomena include changes in pore water content, enrichment or depletion of radiogenic isotopes over time, changes in the flux of cosmogenic radiation, incomplete removal of trapped charge during transport (partial bleaching), and physical disturbance of the grains from bioturbation. For the WHSA Locality 2 samples, the impact of these variables is likely minimal based on the consistency of the measured pore water content between the low-permeability, clay-rich sediment samples; the similarity in radiogenic isotope values between samples; ample Sun exposure during transport prior to burial; and the intact stratigraphy at the sample locations.

Resolution of the White Sands dating controversy requires that there is no combination of physical or chemical processes that could cause the ages of three independent chronologic data sets to converge on a single age range while simultaneously being incorrect or biased. It also requires that paleoenvironmental indicators (pollen assemblages) be consistent with regional records for the period in question. These metrics are met in this study on the basis of the nature of the chronologic techniques we used and the materials to which they were applied. Combined with the previously reported geologic, hydrologic, stratigraphic, chronologic, and climatic evidence (*1*, *30*, *31*), the congruence of the calibrated ^{14}C ages of the aquatic *Ruppia* seeds, the calibrated ^{14}C ages of the terrestrial pollen grains, and the OSL ages support the conclusion that humans were present in North America during the Last Glacial Maximum.

REFERENCES AND NOTES

1. M. R. Bennett *et al*., *Science* **373**, 1528–1531 (2021).
2. P. U. Clark *et al*., *Science* **325**, 710–714 (2009).
3. J. T. Faith, T. A. Surovell, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20641–20645 (2009).
4. D. B. Madsen, L. G. Davis, D. Rhode, C. G. Oviatt, *Science* **375**, eabm4678 (2022).
5. C. G. Oviatt, D. B. Madsen, D. Rhode, L. G. Davis, *Quat. Res.* **111**, 138–147 (2022).
6. C. V. Haynes Jr., *PaleoAmerica* **8**, 95–98 (2021).
7. D. M. Rachal, J. I. Mead, R. Dello-Russo, M. T. Cuba, *Geoarchaeology* **37**, 923–933 (2022).
8. R. S. Bradley, *Paleoclimatology: Reconstructing Climates of the Quaternary* (Academic Press, ed. 3, 2015), pp. 675.
9. T. A. Brown, D. E. Nelson, R. W. Mathewes, J. S. Vogel, J. R. Southon, *Quat. Res.* **32**, 205–212 (1989).
10. T. A. Brown, G. W. Farwell, P. M. Grootes, F. H. Schmidt, *Radiocarbon* **34**, 550–556 (1992).
11. A. Long, O. K. Davis, J. de Lanois, *Radiocarbon* **34**, 557–560 (1992).
12. J. Regnéll, E. Everitt, *Veg. Hist. Archaebot.* **5**, 201 (1996).
13. S. Mensing, J. Southon, *Radiocarbon* **41**, 1–8 (1999).
14. M. R. Kilian, J. Van Der Plicht, B. Van Geel, T. Goslar, *Quat. Int.* **88**, 21–26 (2002).
15. N. Piotrowska, A. Bluszcz, D. Demske, W. Granoszewski, G. Heumann, *Radiocarbon* **46**, 181–187 (2004).
16. W. J. Fletcher *et al*., *Quat. Geochronol.* **39**, 112–123 (2017).
17. C. M. Schiller, C. Whitlock, K. L. Elder, N. A. Iverson, M. B. Abbott, *Radiocarbon* **63**, 321–342 (2021).
18. R. K. Tennant *et al*., *J. Quaternary Sci.* **28**, 229–236 (2013).
19. S. R. H. Zimmerman, T. A. Brown, C. Hassel, J. Heck, *Radiocarbon* **61**, 359–374 (2018).
20. I. Tunno, S. R. H. Zimmerman, T. A. Brown, C. A. Hassel, *Front. Ecol. Evol.* **9**, 668676 (2021).
21. K. Yamada, T. Omori, I. Kitaba, T. Hori, T. Nakagawa, *Quat. Sci. Rev.* **272**, 107236 (2021).
22. Materials and methods are available as supplementary materials.
23. S. A. Hall, in *New Mexico's Ice Ages*, S. G. Lucas, G. S. Morgan, K. E. Zeigler, Eds., *New Mex. Mus. Natur. Hist. Sci. Bull.* **28**, 171–183 (2005).
24. D. J. Donahue, T. W. Linick, A. J. T. Jull, *Radiocarbon* **32**, 135–142 (1990).
25. J. S. Vogel, D. E. Nelson, J. R. Southon, *Radiocarbon* **29**, 323–333 (1987).
26. T. A. Brown, J. R. Southon, *Nucl. Instrum. Methods Phys. Res. B* **123**, 208–213 (1997).
27. A. Murray *et al*., *Nat. Rev. Methods Primers* **1**, 72 (2021).
28. T. Neulieb *et al*., *Radiocarbon* **55**, 1142–1155 (2013).
29. Q. Hua, M. Barbetti, *Radiocarbon* **46**, 1273–1298 (2004).
30. J. S. Pigati *et al*., *Science* **375**, eabm6987 (2022).
31. J. S. Pigati *et al*., *PaleoAmerica* **8**, 99–101 (2022).
32. C. Bronk Ramsey, OxCal 4.4.4 (2022); https://c14.arch.ox.ac.uk/oxcal.html.
33. J. S. Pigati *et al*., Data release for Independent age estimates resolve the controversy of ancient human footprints in White Sands. US Geological Survey data release (2023); https://doi.org/10.5066/P9E36U4B.

ACKNOWLEDGMENTS

Science is a way of knowing, and we acknowledge that there are many ways of knowing. Therefore, we deeply appreciate the perspectives, cultural practices, and oral histories of the tribes and pueblos whose homeland is in southern New Mexico. In addition, we thank White Sands National Park Superintendent M. F. Sauter and the park staff for their support of our research. We also thank P. Henne (US Geological Survey) for preliminary analysis of the pollen assemblages, C. Hassel (Flow Cytometry Resource Facility, Indiana University) for performing the flow cytometry, and T. Brown (Lawrence Livermore National Laboratory) and M. Springer (University of California, Riverside) for helpful discussions. We also thank A. Hudson and M. Berry (both US Geological Survey) and three anonymous reviewers for providing constructive reviews of this manuscript. This study was conducted under National Park Service permits issued to K.B.S. and J.S.P. (WHSA-2021-SCI-0002) and to D.O. (WHSA-2021-SCI-0001). This work was also performed partly under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. This is LLNL-JRNL-844446. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US government. **Funding:** Funding for J.S.P., K.B.S., and J.S.H. was provided by the US Geological Survey's Climate Research and Development Program and by the US Geological Survey–National Park Service Natural Resources Preservation Program. Funding for D.W. and M.R.C. was also provided by the US Geological Survey's Climate Research and Development Program. M.R.B. acknowledges the Arts and Humanities Research Council (AH/X001326/1). D.B. and M.R.B. thank the Western National Parks Association for support. **Author contributions:** J.S.P. and K.B.S. conceptualized the project and completed the field work. J.S.H., D.W., M.R.C., and S.R.H.Z. performed the pollen analysis and radiocarbon laboratory work, and H.J.G. performed the luminescence laboratory work. The original manuscript was drafted by J.S.P. and K.B.S. Editing of the manuscript was done by J.S.P., K.B.S., J.S.H., D.W., M.R.C., S.R.H.Z., H.J.G., V.L.S., D.O., D.B., and MRB. **Competing interests:** The authors declare no competing or conflicting interests. **Data availability:** Downloadable files of the data presented in the supplementary materials (tables S1 and S2) can be found at (*33*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh5007
Materials and Methods
Supplementary Text
Figs. S1 to S10
Tables S1 and S2
References (*34*–*62*)
MDAR Reproducibility Checklist

Submitted 7 March 2023; accepted 10 August 2023
10.1126/science.adh5007

NEUROSCIENCE

Hormone-mediated neural remodeling orchestrates parenting onset during pregnancy

Rachida Ammari[1]†, Francesco Monaca[1]†, Mingran Cao[1], Estelle Nassar[1], Patty Wai[1], Nicholas A. Del Grosso[1], Matthew Lee[1], Neven Borak[1], Deborah Schneider-Luftman[2], Johannes Kohl[1]*

During pregnancy, physiological adaptations prepare the female body for the challenges of motherhood. Becoming a parent also requires behavioral adaptations. Such adaptations can occur as early as during pregnancy, but how pregnancy hormones remodel parenting circuits to instruct preparatory behavioral changes remains unknown. We found that action of estradiol and progesterone on galanin (Gal)–expressing neurons in the mouse medial preoptic area (MPOA) is critical for pregnancy-induced parental behavior. Whereas estradiol silences MPOA$^{\text{Gal}}$ neurons and paradoxically increases their excitability, progesterone permanently rewires this circuit node by promoting dendritic spine formation and recruitment of excitatory synaptic inputs. This MPOA$^{\text{Gal}}$-specific neural remodeling sparsens population activity in vivo and results in persistently stronger, more selective responses to pup stimuli. Pregnancy hormones thus remodel parenting circuits in anticipation of future behavioral need.

Motherhood is associated with pronounced behavioral changes in many species, such as altered feeding routines and increased aggressivity (*1–9*). These adaptations are typically attributed to the hormonal changes associated with giving birth (parturition), which have been hypothesized to activate or prime parental circuits (*10*, *11*). One of the most notable differences between sexually inexperienced (virgin) females and mothers is their infant-directed behaviors: Whereas virgins typically avoid infants or exhibit low levels of parental behavior, mothers are highly parental (*12–14*). Classical studies in rats have found increased maternal responsiveness during pregnancy (*9*, *15*, *16*). This even occurs in females undergoing a caesarean section during mid- or late pregnancy and persists for weeks (*15*, *17–20*). Correspondingly, parental behavior can be elicited in virgin rats by mimicking the hormonal changes of pregnancy (*21–27*), which include drastic rises in the levels of estradiol (E2) and progesterone (P4) (fig. S1A). These observations indicate hormone-mediated, preparatory neural adaptations to infant-directed behavior during pregnancy. However, despite the identification of numerous forms of pregnancy-associated neural plasticity (*1*, *28*, *29*), it remains unknown how pregnancy hormones affect parenting circuits to mediate changes in infant-directed behavior.

Hormone-dependent, long-lasting changes to pup-directed behavior in pregnancy

Whereas virgin female rats and wild house mice typically ignore or attack pups, respectively, virgin female laboratory mice often exhibit hormone-independent, spontaneous parental behavior (*9*). We first investigated when and how pup interactions change during pregnancy. We exposed female mice to pups at regular intervals before, during, and after pregnancy and scored their behaviors (Fig. 1A, *Preg*). Most aspects of parental behavior were affected by pregnancy (Fig. 1, D to G, and fig. S1, B to J), and this was particularly pronounced in late pregnant females [day 18 (D18)]: All D18 females retrieved pups with short latency (D18, 39.7 ± 10.8 s; virgins, 477.9 ± 143.3 s; Fig. 1, D to G), crouched above pups (17.3 ± 3.5% of assay duration), and spent most of their time in the nest (fig. S1, B and C). In addition to individual aspects of parenting, pregnancy affected behavioral sequences: Whereas D18 females performed sequences of retrieval, crouching, nest building, and grooming, virgins engaged in repetitive sniffing–grooming–nest entering episodes (Fig. 1H). The increased parental performance of D18 females could be a result of hormonal effects and/or frequent pup exposure (*30*). We therefore assessed pup interactions in females that were exposed to pups only as virgins and at D18 (Fig. 1B, *Dual*) and in ovariectomized females (Fig. 1C, *OVX*) (*31*). Pup retrieval, crouching, and time in nest differed between virgins (Vir) and D18 females in the *Preg* and *Dual* groups. By contrast, such differences were not present in the *OVX* group over similar time points and are thus primarily affected by pregnancy hormones (Fig. 1, D to G, and fig. S1, B to E). These behaviors were also up-regulated in females that were exposed to pups only once, at D18, which illustrates that the pregnancy-induced onset of parenting does not require any previous pup exposure (fig. S1, N to S). The hormonal milieu of pregnancy thus leads to an onset of specific parental behaviors in mice, and these behavioral changes are maximal in late pregnancy. Most behavioral changes persisted until at least 1 month after parturition (D50) (Fig. 1, D and F, and fig. S1D), when hormone levels have returned to baseline (fig. S1A). These adaptations thus likely result from long-lasting remodeling of the brain by pregnancy hormones.

Hormone action on MPOA$^{\text{Gal}}$ neurons is critical for parenting onset

Parenting is controlled by brain-wide circuits (*32–36*), several elements of which might be affected by hormones. In particular, the medial preoptic area (MPOA)—which is critical for parental behavior—has been shown to be a hormonal target (*37–39*). Parental behavior can be induced in virgin female rats by administration of the hormones E2, P4, prolactin, and oxytocin (*21–26*, *40*, *41*), and global knockout (KO) of their canonical receptors impairs parenting (*42–45*). Because combined systemic administration of E2 and P4 is most effective in triggering parenting onset (*12*), the underlying neural substrates are likely sensitive to both hormones. E2 and P4 can permanently modulate neuronal function through intracellular receptors that act as transcription factors (*46–49*). We focused on the intracellular estrogen receptor 1 (Esr1) and progesterone receptor (PR) because they are critical for parental behavior (*43*, *50–52*) and because the long-lasting nature of pregnancy-induced behavioral changes implicates gene expression–dependent forms of plasticity. Using single-molecule fluorescence in situ hybridization (smFISH) from hypothalamic brain sections of virgins and D18 females, we found that Esr1-PR coexpressing neurons (as well as neurons expressing either receptor) were most enriched in MPOA subregions (fig. S2, A to F). Prolactin receptor, but not oxytocin receptor, expression was similarly enriched in the MPOA (fig. S2, G and H).

To determine whether Esr1 or PR expression in the MPOA is required for parenting onset during pregnancy, mice carrying floxed receptor alleles (materials and methods) (*53*) were injected into the MPOA with an adeno-associated virus (AAV)–expressing Cre recombinase (Fig. 1I). This resulted in local receptor KO, whereas injection of a green fluorescent protein (GFP)–expressing control AAV did not affect receptor expression (fig. S3, A to H). MPOA-specific ablation of either Esr1 or PR had no effect on pup interactions in virgins but completely blocked the pregnancy-induced up-regulation of pup retrieval, crouching, and nest time at D18 (Fig. 1, J to L, and fig. S4, A to C). By contrast, parental behaviors were normally up-regulated at D18 in animals injected with control AAVs (Fig. 1, J to L, and fig. S4, A to D).

Several overlapping populations of MPOA neurons are involved in the control of parenting, with galanin-expressing (MPOA$^{\text{Gal}}$) neurons being critical for this behavior (*33–35*, *54–56*).

[1]State-Dependent Neural Processing Laboratory, The Francis Crick Institute, London NW1 1AT, UK. [2]Bioinformatics and Biostatistics Science Technology Platform, The Francis Crick Institute, London NW1 1AT, UK.
***Corresponding author. Email: jonny.kohl@crick.ac.uk**
†These authors contributed equally to this work.

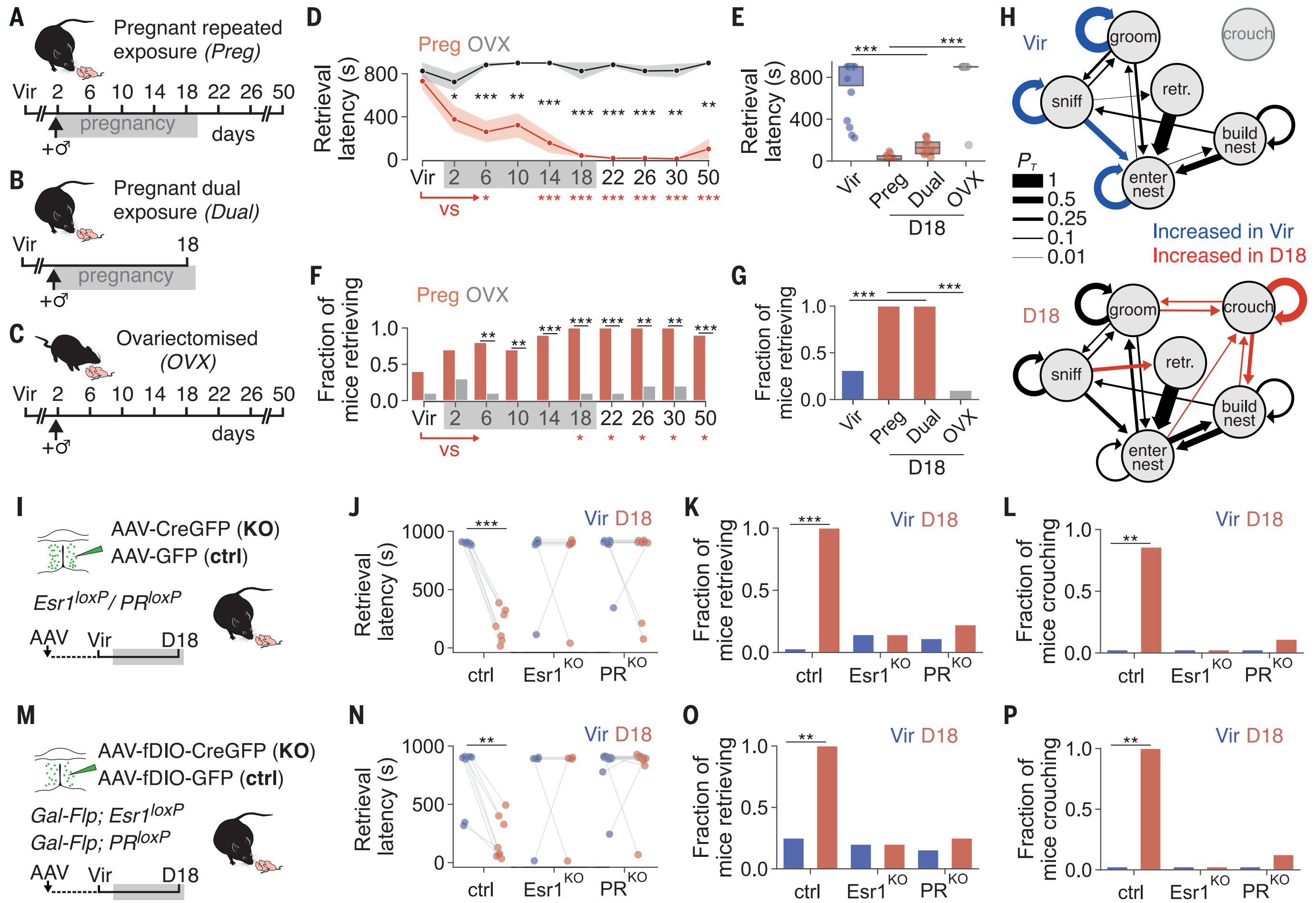


Fig. 1. Hormone action on MPOA^{Gal} neurons is critical for pregnancy-induced onset of parental behavior. (**A** to **C**) Testing pup-directed behavior in repeatedly pup-exposed pregnant females (*Preg*, *n* = 10) (A), pregnant females exposed to pups twice (*Dual*, *n* = 9) (B), and repeatedly pup-exposed ovariectomized females (*OVX*, *n* = 10) (C). Day of pregnancy [(A) and (B)] or relative to pairing with male (C) are shown. (**D** and **F**) Parental behaviors in *Preg* group. Within-group [*Preg*, virgin (Vir) versus each subsequent time point; red asterisks] and between-group (*Preg* versus *OVX*; black asterisks) comparisons are shown. (**E** and **G**) Comparison of Vir and D18 time points across groups. Note that virgins from *Preg* and *Dual* groups are pooled (fig. S1M). (**H**) Behavioral state transition diagrams for Vir and D18 females (*Preg*, *n* = 10). Average transition probabilities (P_T) between behaviors are shown, and differences between Vir and D18 females are highlighted if $P < 0.05$ (*U* test; see materials and methods). retr., retrieval. (**I**) AAV-mediated ablation (KO) of Esr1 or PR in MPOA and control (ctrl). (**J** to **L**) Effects of MPOA-wide KO of Esr1 or PR on pup-directed behaviors (*n* = 7, 8, and 9 mice). (**M**) KO of Esr1 or PR in MPOA^{Gal} neurons. (**N** to **P**) Effects of MPOA^{Gal}-specific KO of Esr1 or PR on pup-directed behaviors (*n* = 8, 5, and 13 mice). Statistics by Kaplan-Meier survival analysis with log rank test in (D), (E), (J), and (N) and by Fisher's exact test with Benjamini-Hochberg adjustment for multiple comparisons in (F), (G), (K), (L), (O), and (P). Shaded area in (D) is SEM. ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$.

MPOA^{Gal} neurons, most of which express Esr1 and PR (fig. S3, J and K), constitute ~20% of MPOA neurons (*35*). To determine whether hormonal sensitivity of this subpopulation is necessary for pregnancy-induced behavioral adaptations, we made a knock-in mouse line expressing Flp recombinase in Gal neurons (fig. S3I and materials and methods) and crossed this allele into mice with floxed receptor genes. AAV-mediated ablation of either Esr1 or PR in MPOA^{Gal} neurons fully recapitulated the effects observed after MPOA-wide receptor KO (Fig. 1, M to P, and fig. S4, E to H). By contrast, pup contact latency—a parameter not modulated by pregnancy—was not affected by this manipulation (fig. S4I). The parental behaviors of these receptor-ablated animals remained impaired even after giving birth (D22; fig. S5), which indicates that the lack of hormone-mediated behavioral preparations during pregnancy cannot be compensated for by the subsequent endocrine events of parturition. Direct action of E2 and P4 on MPOA^{Gal} neurons through their intracellular hormone receptors is therefore necessary for pregnancy-mediated increases in parental behavior.

Long-lasting hormonal remodeling of MPOA^{Gal} neurons during pregnancy

We next investigated how pregnancy affects MPOA^{Gal} neurons and performed patch clamp recordings in brain slices from virgins and D18 females (Fig. 2A, top). Recorded neurons were neurobiotin-filled and reconstructed to assess morphological changes. MPOA^{Gal} neurons exhibited lower baseline firing and resting membrane potential in late pregnancy (Fig. 2, B and C, and supplementary text), with a significantly higher proportion of silent neurons at D18 (fig. S6, A and B). This silencing was abolished by Tertiapin-Q (TQ) and might thus be mediated by GIRK channels (Fig. 2C). At the same time, MPOA^{Gal} neurons were more excitable at D18 and less frequently exhibited depolarization block (Fig. 2, D and E, and fig. S6K). We observed a reduction in action potential half-width at D18 (fig. S6E), which hints at increased function of delayed rectifier K^+ channels that repolarize neurons to permit sustained firing (*57*). These effects on neuronal membrane

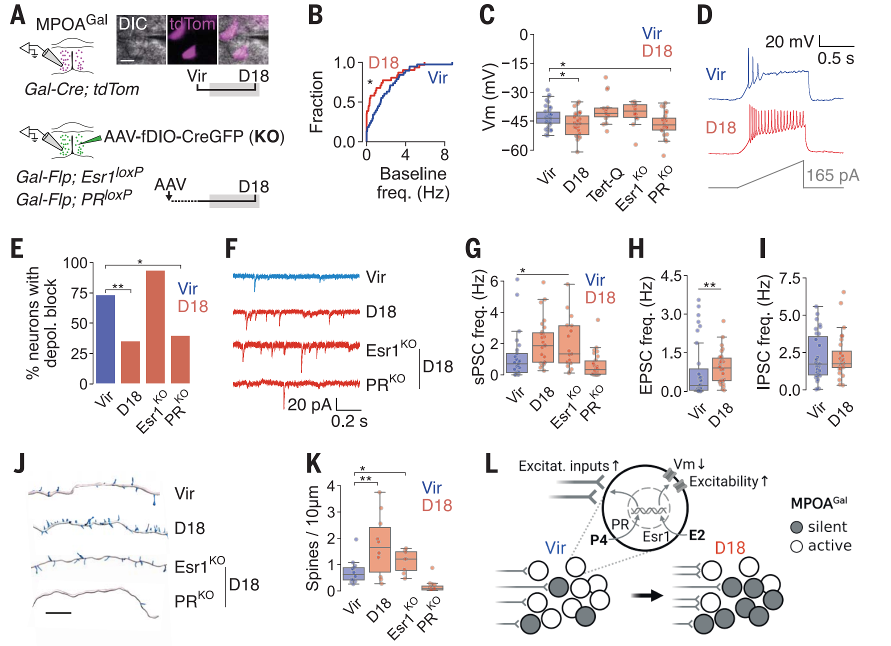


Fig. 2. Hormonal remodeling of MPOA^{Gal} neurons. (**A**) Whole-cell recordings from wild-type (top) and receptor-deleted (KO) (bottom) MPOA^{Gal} neurons. (**B**) Cumulative distribution of baseline firing frequency (Vir and D18; 33 and 21 cells from n = 15 and 7 mice). (**C**) Resting membrane potential of control and receptor-deleted MPOA^{Gal} neurons and recordings in presence of GIRK channel blocker Tertiapin-Q (Tert-Q) (38, 32, 15, 18, and 26 cells from n = 15, 9, 3, 3, and 5 mice). (**D**) Example current clamp recording traces of cells with (Vir) and without (D18) depolarization block. (**E**) Percentage of neurons exhibiting depolarization block (34, 30, 18, and 25 cells from n = 15, 8, 3, and 5 mice). (**F**) Example voltage clamp recording traces with sPSCs. (**G**) sPSC frequency (21, 23, 18, and 26 cells from n = 9, 6, 3, and 5 mice). (**H** and **I**) Excitatory postsynaptic current (EPSC) (Vir and D18; 31 and 30 cells from n = 5 and 4 mice) (H) and inhibitory postsynaptic current (IPSC) (31 and 28 cells from n = 5 and 4 mice) (I) frequency. (**J**) Dendritic segments of MPOA^{Gal} neurons with spines. (**K**) Spine density (14, 10, 8, and 15 cells from n = 10, 4, 3, and 4 mice). (**L**) Summary scheme for hormonal remodeling of MPOA^{Gal} neurons. Statsitics by U test in (B); one-way analysis of variance (ANOVA) with Dunnett's post hoc test in (C), (G), and (K); Fisher's exact test with Benjamini-Hochberg adjustment in (E); and K-S test in (H) and (I). Scale bars, 20 µm (A) and 10 µm (J). ***P < 0.001; **P < 0.01; *P < 0.05.

properties were already apparent in midpregnancy (D10; fig. S6, B to E) and were linked because 80% of silent MPOA^{Gal} neurons also did not exhibit depolarization block at D18 (fig. S6O). MPOA^{Gal} silencing was not a result of increased inhibitory synaptic inputs: Although these neurons received more spontaneous postsynaptic currents (sPSCs) at D18, this was because of an increase in excitatory inputs (Fig. 2, F to I, and fig. S6M), which predominantly targeted spontaneously active—i.e., nonsilenced—neurons (fig. S6N). Correspondingly, MPOA^{Gal} neurons had more dendritic spines at D18 (Fig. 2, J and K). This remodeling of synaptic inputs was also already detectable at D10 (fig. S6, I and J). We did not observe changes to sPSC amplitude and dynamics (fig. S6L) or to dendritic complexity and somatic volume (fig. S6, P and Q) (*58*). Pregnancy therefore reduces the baseline activity of MPOA^{Gal} neurons while increasing their excitability and promoting the recruitment of excitatory synaptic inputs. Pregnancy did not have equivalent effects on Gal-negative MPOA neurons, which highlights the specificity of MPOA^{Gal} neuronal remodeling (fig. S7).

To address whether these biophysical and morphological changes were the result of direct hormonal action, we recorded from MPOA^{Gal} neurons in which Esr1 or PR were deleted (Fig. 2A, bottom). Ablation of these receptors returned specific, nonoverlapping aspects of D18 neuronal physiology to a virgin-like state: Esr1 deletion specifically prevented pregnancy-induced silencing and changes to excitability (Fig. 2, C and E) but did not affect synaptic inputs and spine density (Fig. 2, F, G, J, and K). By contrast, PR deletion selectively abolished the up-regulation of synaptic inputs and spine density (Fig. 2, G and K) without affecting membrane properties (Fig. 2, C and E). Transduction with a control AAV had no effect (fig. S6, R to W). E2 and P4 therefore control discrete aspects of pregnancy-induced plasticity in MPOA^{Gal} neurons: Whereas E2 tunes membrane potential and intrinsic excitability, P4 mediates the recruitment of additional excitatory synaptic inputs (Fig. 2L). To assess how long-lasting these changes were, we recorded from MPOA^{Gal} neurons in mothers shortly after parturition (D22) and at D50, when pregnancy- and parturition-associated hormone levels have returned to baseline (fig. S8A). MPOA^{Gal} resting membrane potential and firing frequency remained reduced at D22 and only returned to virgin-like levels at D50 (fig. S8, B and C), whereas neuronal excitability reverted immediately after parturition (fig. S8, E and F). By contrast, synaptic inputs and spine density showed a long-lasting up-regulation (fig. S8, G and H). Similar to the lasting behavioral effects of receptor ablation, its physiological effects persisted in mothers (fig. S8, I to K). These observations suggest that pregnancy hormones permanently alter the circuit integration of MPOA^{Gal} neurons, thereby providing a cellular substrate for the long-lasting behavioral effects of pregnancy.

Reorganization of MPOA^{Gal} neuronal and neural population activity during pregnancy

We next investigated the effects of pregnancy on MPOA^{Gal} neural activity in vivo. We performed longitudinal, cellular-resolution calcium imaging from MPOA^{Gal} neurons in females exposed to pups and a set of social and nonsocial stimuli (Fig. 3, A to C, and fig. S9A) (*59*). Consistent with the silencing observed in our slice physiology recordings, the number of detectable (nonsilent) MPOA^{Gal} neurons was significantly reduced at D18 in vivo (Fig. 3, D and E). This reduction was not the result of a decline in the number of GCaMP-expressing MPOA^{Gal} neurons over time or increased calcium buffering by rising GCaMP levels because it was reversible (Fig. 3E) and did not occur in virgin females recorded at identical time points (fig. S9B). It also did not result from potential shifts in the recording plane because we observed this effect when imaging ex vivo (fig. S9, E and F). Finally, the number of detected neurons was not significantly decreased in MPOA-wide recordings (fig. S9, C and D). Pregnancy-induced silencing therefore preferentially occurs in MPOA^{Gal} neurons, consistent with our electrophysiological findings in brain slices (fig. S7, A to E).

The fraction of neurons activated during pup retrieval and pup grooming decreased at D18 (Fig. 3, F and G, and fig. S10A). By contrast, similar fractions of MPOA^{Gal} neurons were active during pup sniffing in virgins and at D18, but their responses occurred with shorter latency at D18 and D50 (Fig. 3, I to K), which indicates a higher excitability of MPOA^{Gal} neurons to pup stimuli during and after pregnancy. Pregnancy therefore sparsens MPOA^{Gal} population activity during parental actions and makes these neurons more excitable to pup stimuli. The baseline activity of individual

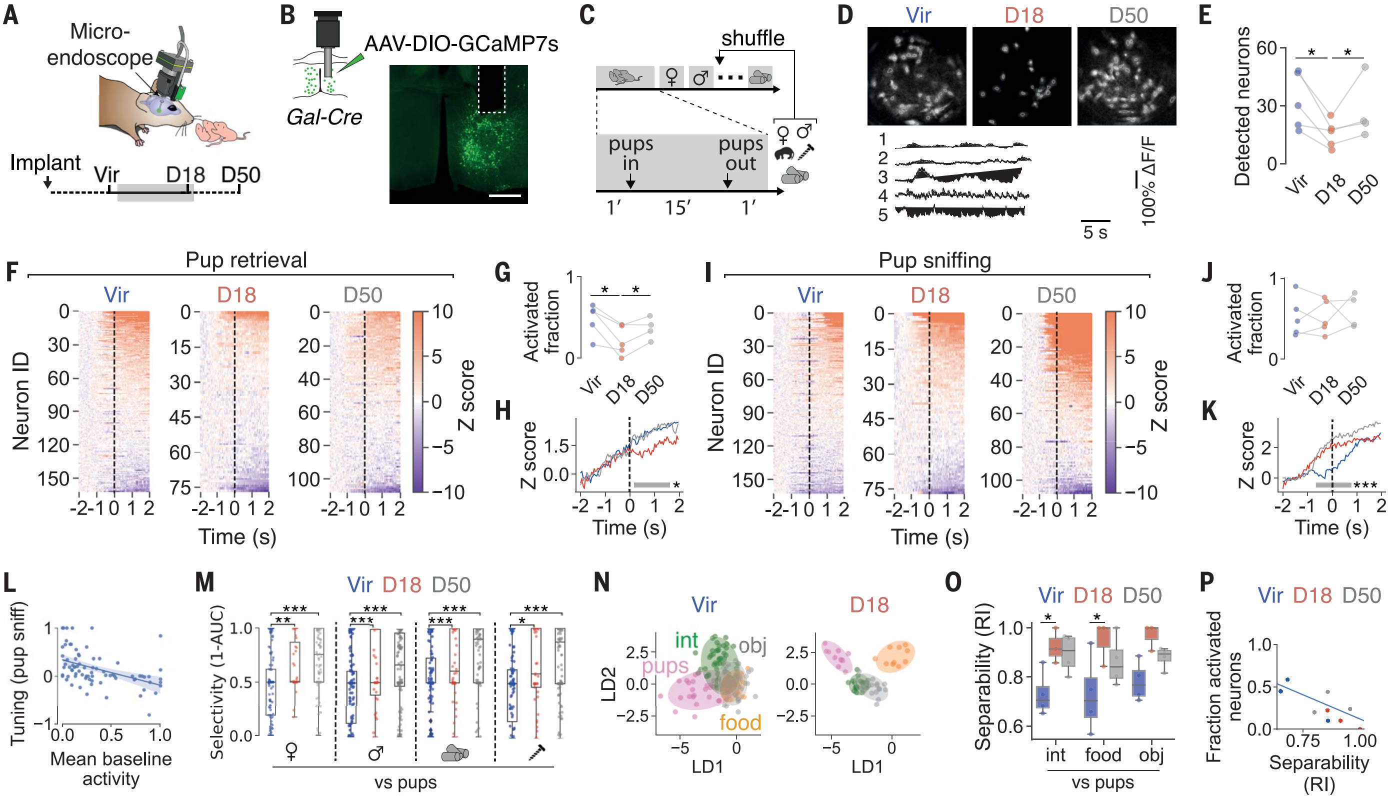


Fig. 3. Reorganization of MPOA^{Gal} population activity during pregnancy. (**A**) Recording setup for miniature microscope recordings. (**B**) *Gal-Cre* animals were injected into the MPOA with AAV-FLEx-GCaMP7s and implanted with a GRIN lens. GCaMP7s expression and GRIN lens position are shown. (**C**) Experimental design (see materials and methods). (**D**) Sample recording frames with detected neurons and example activity traces from a virgin. (**E**) Number of detected (nonsilent) neurons per animal (n = 5 mice). (**F** and **I**) Temporal profile of MPOA^{Gal} responses during pup retrieval (F) or sniffing (I) in virgins, at D18, and at D50 (162, 77, and 93 neurons from n = 5 mice). Dashed lines indicate action onset. Order was based on hierarchical clustering sorted by mean cluster response onset. (**G** and **J**) Fraction of neurons with positive evoked response during pup retrieval (G) or sniffing (J) (n = 5 mice). (**H** and **K**) Averaged Z score for neurons activated during pup retrieval in virgins, at D18, and at D50 (115, 41, and 63 neurons from 5, 5, and 4 mice) (H) or sniffing (122, 51, and 86 neurons from 5, 5, and 4 mice) (K). Statistics by two-way ANOVA with Tukey post hoc test, and gray bars indicate periods of significant difference for Vir versus D18 and Vir versus D50. (**L**) Correlation between normalized tuning index for responses to pup sniffing and normalized mean baseline activities at D18 (coefficient of determination r^2 = 0.202; P < 0.001). (**M**) Selectivity of chemoinvestigation-associated responses for indicated stimulus pairs at Vir, D18, and D50 (142, 35, and 108 cells from n = 4, 3, and 4 mice) compared with pups. A selectivity score of 1 means the neuron is only activated during pup sniffing, a score of 0 means selective activation during sniffing of other stimulus, and a score of 0.5 equals a nonselective response (see materials and methods). (**N**) Example MPOA^{Gal} neuronal activity at Vir and D18 during object investigation in LDA space (int, intruder; obj, object). Temporal bins were used as features. Ellipsoids represent 95% confidence area of neuronal activity to each stimulus. (**O**) Separability of indicated stimulus combinations by the MPOA^{Gal} population [Rand Index (RI), n = 4, 3, and 4 mice]. (**P**) Correlation between separability (pup versus intruder) and activated fraction of neurons during pup retrieval (r^2 = 0.56; P < 0.001). Statistics by paired t tests in (E), (G), and (J); mixed linear model with mouse ID as group in (M); linear regression in (L) and (P); and unpaired t tests in (O). Scale bar in (B), 500 μm. ***P < 0.001; **P < 0.01; *P < 0.05.

MPOA^{Gal} neurons was negatively correlated with their tuning to pup stimuli at D18, thereby linking neuronal silencing to stronger pup-evoked responses (Fig. 3L). To address pregnancy-induced differences in how MPOA^{Gal} neurons represent pup stimuli, we examined their activity patterns during chemoinvestigation of pups and other stimuli (Fig. 3C). MPOA^{Gal} neuronal stimulus selectivity for, and response strength to, pups increased in late pregnancy (Fig. 3M and fig. S11, A to E). Similarly, although linear discriminant analysis (LDA) could not separate pup representations in MPOA^{Gal} neurons well from those of other stimuli in virgins, separability of pup representations from those of other stimuli was enhanced at D18 (Fig. 3, N and O). Stimulus separability was positively correlated with population sparsening, thereby linking improved pup representations with effective encoding of parental actions (Fig. 3P). At D50, the numbers of spontaneously active and retrieval-activated neurons had largely returned to virgin levels (Fig. 3, E and G), mirroring our findings in brain slices (fig. S8, B to D). By contrast, pup stimulus selectivity and separability showed a long-lasting increase (Fig. 3, M to O). These findings demonstrate that pregnancy leads to a pronounced sparsening of spontaneous and parenting-associated activity in MPOA^{Gal} neurons and to increased selectivity for infant stimuli.

Discussion

Considerable progress has been made in uncovering the functional architecture of parenting circuits (*9*, *32*–*36*), but little is known about how hormones alter these circuits to ensure state-dependent behavioral flexibility. We discovered that pregnancy hormone action on MPOA^{Gal} neurons—a hub in parenting circuits—is critical to instruct a preparatory change in infant-directed behavior. The ovarian hormones E2 and P4 each control distinct aspects of pregnancy-induced neural remodeling: Whereas E2 transiently silences MPOA^{Gal} neurons and increases their excitability, P4 permanently remodels this circuit element by recruiting synaptic inputs. This results in sparsened population

activity during parental behavior and in potentiated, more-selective responses to pup stimuli. We propose that the resulting increase in signal-to-noise ratio, both in individual neurons and at the population level, enables more efficient encoding of parental motor actions by MPOA^{Gal} neurons. Population sparsening through silencing might contribute to setting up the circuit for efficient parental behavior by selectively recruiting inputs onto active MPOA^{Gal} neurons during pregnancy. Once rewired, this circuit could then drive robust parenting in response to pup cues, whereas release from silencing during the postpartum period might allow for recruitment of these neurons during nonparental social interactions.

The long-lasting, P4-mediated remodeling of MPOA^{Gal} synaptic inputs provides a cellular correlate for the long-lasting behavioral changes that we observe. Although parturition-associated hormonal changes and subsequent maternal experience cannot compensate for lack of hormonal remodeling during pregnancy, these factors might normally contribute to long-term enhancement of maternal behavior (maternal memory) (*9*). Repeated and/or prolonged cohousing of virgins with pups results in elevated levels of parenting through sensitization (*60*, *61*), and parental care can be socially transmitted by mothers (*62*). It is unclear whether these paradigms result in similar neuronal changes. Ablating MPOA^{Gal} neurons or making them hormone-insensitive both abolish pup retrieval, but optogenetic activation of these neurons elicits pup grooming in virgins (*35*). Although it remains unknown which neuronal ensembles are recruited by artificial, acute stimulation, they seem to differ from the sparse populations that drive robust parenting in late pregnancy.

E2 silences MPOA^{Gal} neurons beyond parturition, presumably by up-regulating GIRK channel expression, whereas the more transient increases in excitability are likely a result of potentiated function of delayed rectifier K^+ channels (*57*). The identity of the additional excitatory inputs recruited by P4 remains unknown. They might constitute long-range afferents conveying pup sensory information because most local MPOA neurons are GABAergic [γ-aminobutyric acid (GABA)] (*54*). Future work will characterize the identity and functional role of the cellular pathways targeted by Esr1 and PR. Coexpression of these receptors is not unique to Gal-expressing MPOA neurons (fig. S2C). We hypothesize that permissive chromatin states in these neurons allow for cell type–specific, hormonally induced target gene expression. MPOA^{Gal} neurons form molecularly distinct subpopulations (*54*) that might be differentially affected by pregnancy hormones. We also expect MPOA^{Gal} and other neurons in parenting circuits to be sensitive to additional pregnancy hormones, such as prolactin, placental lactogens, allopregnanolone, and oxytocin. Prolactin for instance, which acts on the MPOA in early pregnancy to reduce physical activity (*63*), might also contribute to early changes in pup-directed behavior (Fig. 1, D and F).

Unlike laboratory mice, the majority of wild virgin female mice exhibit infanticide (*64*). Our work provides mechanisms through which hormones might act in parental circuits of wild mice and other species that critically depend on endocrine changes for the onset of short-latency maternal behavior, such as rats, rabbits, and sheep (*9*). The neural activity changes observed in this work—i.e., population sparsening and increased stimulus selectivity and discriminability—are reminiscent of changes that occur during critical periods in the developing brain (*65*). Our work therefore suggests that pregnancy hormones open a window of adult plasticity during which neural remodeling orchestrates behavioral adaptations for the future challenges of motherhood.

REFERENCES AND NOTES

1. P. J. Brunton, J. A. Russell, *Nat. Rev. Neurosci.* **9**, 11–25 (2008).
2. E. N. Makarova, E. D. Kochubei, N. M. Bazhan, *Neurosci. Behav. Physiol.* **40**, 263–267 (2010).
3. M. A. Mann, C. Konen, B. Svare, *Horm. Behav.* **18**, 140–160 (1984).
4. J. A. Russell, A. J. Douglas, C. D. Ingram, *Prog. Brain Res.* **133**, 1–38 (2001).
5. M.-C. Buhot-Averseng, J. Goyens, *Aggress. Behav.* **7**, 111–121 (1981).
6. R. Haddad-Tóvolli *et al.*, *Nat. Metab.* **4**, 424–434 (2022).
7. R. D. Lisk, *Anim. Behav.* **19**, 606–610 (1971).
8. T. Topilko *et al.*, *Neuron* **110**, 1385–1399.e8 (2022).
9. M. Numan, *The Parental Brain: Mechanisms, Development, and Evolution* (Oxford Univ. Press, 2020).
10. M. Numan, B. Woodside, *Behav. Neurosci.* **124**, 715–741 (2010).
11. M. Numan, L. J. Young, *Horm. Behav.* **77**, 98–112 (2016).
12. J. S. Rosenblatt, A. D. Mayer, A. L. Giordano, *Psychoneuroendocrinology* **13**, 29–46 (1988).
13. A. S. Fleming, J. S. Rosenblatt, *J. Comp. Physiol. Psychol.* **86**, 957–972 (1974).
14. M. Numan, T. R. Insel, *The Neurobiology of Parental Behavior*, vol. 1 of *Hormones, Brain, and Behavior*, G. F. Ball, J. Balthazart, R. J. Nelson, Eds. (Springer, 2003).
15. J. S. Rosenblatt, *Am. J. Orthopsychiatry* **39**, 36–56 (1969).
16. B. M. Slotnick, M. L. Carpenter, R. Fusco, *Horm. Behav.* **4**, 53–59 (1973).
17. R. S. Bridges, *Physiol. Behav.* **18**, 487–490 (1977).
18. J. Labriola, *Proc. Soc. Exp. Biol. Med.* **83**, 556–557 (1953).
19. H. Moltz, D. Robbins, M. Parks, *J. Comp. Physiol. Psychol.* **61**, 455–460 (1966).
20. B. P. Wiesner, N. M. Sheard, *Maternal Behavior in the Rat* (Oliver and Boyd, 1933).
21. R. S. Bridges, M. Numan, P. M. Ronsheim, P. E. Mann, C. E. Lupini, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8003–8007 (1990).
22. A. D. Mayer, H. B. Ahdieh, J. S. Rosenblatt, *Horm. Behav.* **24**, 152–173 (1990).
23. H. Moltz, M. Lubin, M. Leon, M. Numan, *Physiol. Behav.* **5**, 1373–1377 (1970).
24. C. A. Pedersen, J. A. Ascher, Y. L. Monroe, A. J. Prange Jr., *Science* **216**, 648–650 (1982).
25. O. Riddle, E. L. Lahr, R. W. Bates, *Exp. Biol. Med.* **32**, 730–734 (1935).
26. J. S. Rosenblatt, A. Olufowobi, H. I. Siegel, *Horm. Behav.* **33**, 104–114 (1998).
27. J. Terkel, J. S. Rosenblatt, *J. Comp. Physiol. Psychol.* **65**, 479–482 (1968).
28. K. M. Hillerer, V. R. Jacobs, T. Fischer, L. Aigner, *Neural Plast.* **2014**, 574159 (2014).
29. N. Zilkha, N. Scott, T. Kimchi, *Annu. Rev. Neurosci.* **40**, 273–305 (2017).
30. B. J. Marlin, M. Mitre, J. A. D'amour, M. V. Chao, R. C. Froemke, *Nature* **520**, 499–504 (2015).
31. P. M. Wise, A. Ratner, *Neuroendocrinology* **30**, 15–19 (1980).
32. P. B. Chen *et al.*, *Cell* **176**, 1206–1221.e18 (2019).
33. Y.-Y. Fang, T. Yamaguchi, S. C. Song, N. X. Tritsch, D. Lin, *Neuron* **98**, 192–207.e10 (2018).
34. J. Kohl *et al.*, *Nature* **556**, 326–331 (2018).
35. Z. Wu, A. E. Autry, J. F. Bergan, M. Watabe-Uchida, C. G. Dulac, *Nature* **509**, 325–330 (2014).
36. S. Stagkourakis *et al.*, *Cell* **182**, 960–975.e15 (2020).
37. M. Numan, J. S. Rosenblatt, B. R. Komisaruk, *J. Comp. Physiol. Psychol.* **91**, 146–164 (1977).
38. S. E. Fahrbach, D. W. Pfaff, *Horm. Behav.* **20**, 354–363 (1986).
39. Y. Tsuneoka *et al.*, *J. Comp. Neurol.* **521**, 1633–1663 (2013).
40. R. S. Bridges, P. M. Ronsheim, *Endocrinology* **126**, 837–848 (1990).
41. A. D. Mayer, M. A. Monroy, J. S. Rosenblatt, *Horm. Behav.* **24**, 342–364 (1990).
42. B. K. Lucas, C. J. Ormandy, N. Binart, R. S. Bridges, P. A. Kelly, *Endocrinology* **139**, 4102–4107 (1998).
43. S. Ogawa *et al.*, *Endocrinology* **139**, 5070–5081 (1998).
44. M. E. Rich, E. J. deCárdenas, H.-J. Lee, H. K. Caldwell, *PLOS ONE* **9**, e98839 (2014).
45. Y. Takayanagi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16096–16101 (2005).
46. E. R. Levin, *Trends Endocrinol. Metab.* **20**, 477–482 (2009).
47. N. Z. Lu *et al.*, *Pharmacol. Rev.* **58**, 782–797 (2006).
48. P. Thomas, Y. Pang, *Neuroendocrinology* **96**, 162–171 (2012).
49. C. D. Toran-Allerand *et al.*, *J. Neurosci.* **22**, 8391–8401 (2002).
50. R. D. Brinton *et al.*, *Front. Neuroendocrinol.* **29**, 313–339 (2008).
51. A. C. Ribeiro *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16324–16329 (2012).
52. M. Numan *et al.*, *Brain Res.* **830**, 358–371 (1999).
53. P. Antonson, Y. Omoto, P. Humire, J.-Å. Gustafsson, *Biochem. Biophys. Res. Commun.* **424**, 710–716 (2012).
54. J. R. Moffitt *et al.*, *Science* **362**, eaau5324 (2018).
55. Y.-C. Wei *et al.*, *Nat. Commun.* **9**, 279 (2018).
56. C. Yoshihara *et al.*, *Cell Rep.* **35**, 109204 (2021).
57. J. A. McHenry *et al.*, *Nat. Neurosci.* **20**, 449–458 (2017).
58. L. Keyser-Marcus *et al.*, *Brain Res. Bull.* **55**, 737–745 (2001).
59. K. K. Ghosh *et al.*, *Nat. Methods* **8**, 871–878 (2011).
60. J. S. Rosenblatt, *Science* **156**, 1512–1514 (1967).
61. D. S. Stolzenberg, E. F. Rissman, *J. Neuroendocrinol.* **23**, 345–354 (2011).
62. I. Carcea *et al.*, *Nature* **596**, 553–557 (2021).
63. S. R. Ladyman, K. M. Carter, M. L. Gillett, Z. K. Aung, D. R. Grattan, *eLife* **10**, e62260 (2021).
64. M. M. McCarthy, F. S. vom Saal, *Physiol. Behav.* **35**, 843–849 (1985).
65. G. B. Smith *et al.*, *Nat. Neurosci.* **18**, 252–261 (2015).
66. J. Kohl, AmmariMonaca2023, Figshare (The Francis Crick Institute, 2023); https://doi.org/10.25418/crick.c.6706164.v1.
67. J. Kohl, M. Cao, F. Monaca, Hormone-mediated neural remodeling orchestrates parenting onset during pregnancy, version v1, Zenodo (2023); https://doi.org/10.5281/zenodo.8222741.

ACKNOWLEDGMENTS

We thank M. Strom for advice on and production of viruses and A. Schaefer, F. Iacaruso, P. Znamenskiy, F. Guillemot, J. Briscoe, Y. Isogai, L. Newman, and all members of the State-Dependent Neural Processing Laboratory for discussions and comments on the manuscript. We thank the animal facilities at the Francis Crick Institute for animal care and technical assistance as well as the Light Microscopy, Experimental Histopathology, Genetic Modification Services, and Bioinformatics science technology platforms; the Making Lab; and the Mechanical Workshop. **Funding:** This study received support from the Francis Crick Institute, core funding FC001153. The Francis Crick Institute receives its funding from Cancer Research UK, the UK Medical Research Council, and the Wellcome Trust (J.K.). This study was also supported by European Research Council starting grant ERC-2019-STG847873 (J.K.), NARSAD Young Investigator Award BB/V016946/1 (J.K.), and a Boehringer Ingelheim PhD fellowship (M.C.). **Author contributions:** Conceptualization: J.K., R.A., F.M., and M.C. Methodology: R.A., F.M., M.C., P.W., and J.K. Investigation: R.A., F.M., M.C., P.W., M.L., D.S.-L., and J.K. Formal analysis: R.A., F.M., M.C., E.N., N.A.D.G.,

N.B., D.S.-L., and J.K. Funding acquisition: J.K. and M.C. Supervision: J.K. Writing – original draft: J.K. Writing – review & editing: J.K., F.M., R.A., and M.C. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials or are deposited at Figshare (*66*). Analysis code is available online (https://github.com/FrancisCrickInstitute/AmmariMonaca2023) (*67*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adi0576
Materials and Methods
Supplementary Text
Figs. S1 to S11
References (*68*–*78*)
MDAR Reproducibility Checklist
Movie S1

Submitted 11 May 2023; accepted 9 August 2023
10.1126/science.adi0576

2D MATERIALS

A quantum ruler for orbital magnetism in moiré quantum matter

M. R. Slot[1,2]†, Y. Maximenko[1,3]†, P. M. Haney[1], S. Kim[1,4], D. T. Walkup[1], E. Strelcov[1,3], Son T. Le[1], E. M. Shih[1,3], D. Yildiz[1,4], S. R. Blankenship[1], K. Watanabe[5], T. Taniguchi[6], Y. Barlas[7], N. B. Zhitenev[1], F. Ghahari[8]*, J. A. Stroscio[1]*

For almost a century, magnetic oscillations have been a powerful "quantum ruler" for measuring Fermi surface topology. In this study, we used Landau-level spectroscopy to unravel the energy-resolved valley-contrasting orbital magnetism and large orbital magnetic susceptibility that contribute to the energies of Landau levels of twisted double-bilayer graphene. These orbital magnetism effects led to substantial deviations from the standard Onsager relation, which manifested as a breakdown in scaling of Landau-level orbits. These substantial magnetic responses emerged from the nontrivial quantum geometry of the electronic structure and the large length scale of the moiré lattice potential. Going beyond traditional measurements, Landau-level spectroscopy performed with a scanning tunneling microscope offers a complete quantum ruler that resolves the full energy dependence of orbital magnetic properties in moiré quantum matter.

Moiré quantum matter (MQM) systems (*1*) consist of stacked and twisted layers of van der Waals materials and have emerged as versatile condensed-matter quantum simulators (*2*). The twist angle, choice of material, number of layers, and the application of electric and magnetic fields provide a vast arena for realizing quantum phases that result from the interplay between electron correlations and topology. The discovery of flat electronic bands that host superconductivity and correlated insulating states in magic-angle twisted bilayer graphene (MATBG) inspired rapid exploration of the parameter space (*3*–*5*). The most recent endeavors have focused on heterostructures of three, four, and five layers of alternating twisted graphene monolayers (*6*–*11*).

The moiré systems of MATBG and related heterostructures support topological bands with nonzero Chern number (*4*, *12*), which is derived from the Berry curvature of the Bloch wave functions (*13*). Berry curvature is intimately related to orbital magnetization, and indeed, previous work on MATBG and related systems has observed orbital magnetic order, valley Hall, and quantum anomalous Hall effects (*14*–*16*). Berry curvature and the related quantum metric also provide significant contributions to the orbital magnetic susceptibility (*17*, *18*). The valley-contrasting orbital magnetism and orbital magnetic susceptibility play a central role in the material response to applied magnetic fields. Landau-level (LL) spectroscopy is a well-established tool to experimentally deduce the zero-field properties of the band structure, and in this work, we expand its application to extracting higher-order magnetic response functions.

Semiclassically, LLs are described by the Onsager relation, relating the extremal cross-sectional area of the Fermi surface to the period of oscillations in the de Haas–van Alphen effect (*19*). The orbital magnetism and orbital magnetic susceptibility give rise to energy-dependent first- and second-order corrections with magnetic field in the Onsager relation (*20*), which was recently formulated as a general expansion in higher-order response functions (*21*). To determine these effects and their full energy dependence, high-resolution measurements of the LLs in the electronic bands are required: a "quantum ruler" probing the evolution of LLs at all energies as a function of displacement field and magnetic field.

In this work, we used LL measurements as a quantum ruler for twisted double-bilayer graphene (TDBG) to quantitatively determine the tunable electronic structure, energy-dependent valley-contrasting orbital magnetism, and orbital magnetic susceptibility. TDBG is MQM with highly reconfigurable bands; the Bernal bilayer constituents have an electrostatically tunable band structure in themselves (*22*–*31*). Flat bands exist over a wide range of small twist angles (θ), from about 0.8° to 1.5°. Previous transport and scanning tunneling microscopy (STM) measurements have focused on correlated states, which form at various partial fillings (*22*–*30*), or density wave states at larger twist angles of θ ≈ 2.4° (*31*). Our intermediate twist angle of 1.74° gives rise to moiré minibands that are narrow but not extremely flat. This slightly wider bandwidth allows LLs to be resolved over the entire band structure, which has not been observed in previous local probe measurements (*29*).

[1]Physical Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA. [2]Department of Physics, Georgetown University, Washington, DC 20007, USA. [3]Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA. [4]Joint Quantum Institute, Department of Physics, University of Maryland, College Park, MD 20742, USA. [5]Research Center for Functional Materials, National Institute for Materials Science, 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan. [6]International Center for Materials Nanoarchitectonics, National Institute for Materials Science, 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan. [7]Department of Physics, University of Nevada, Reno, NV 89557, USA. [8]Department of Physics and Astronomy, George Mason University, Fairfax, VA 22030, USA.
*Corresponding author. Email: fghahari@gmu.edu (F.G.); joseph.stroscio@nist.gov (J.A.S.)
†These authors contributed equally to this work.

Tuning the electronic structure of TDBG

Figure 1A shows the simulated band structure of TDBG at θ = 1.75° with the Bistritzer–MacDonald (BM) continuum model (*3*). The moiré periodic potential leads to narrow moiré minibands, composed of two low-energy bands—the valence V1 and conduction C1 bands—and higher remote energy bands—V2 and C2. The narrow bands V1 and C1 have a bandwidth of about 50 meV and are isolated from other bands (Fig. 1B). The TDBG system displays a large tunability in its electronic structure (*12*, *32*–*39*). A perpendicular electric field induces a potential difference between the layers of TDBG, drastically altering the V1 band structure. The V1 band is relatively flat at zero displacement field (D) (Fig. 1B) and develops pronounced electron and hole pockets at $D \neq 0$ (Fig. 1C and fig. S8). In the next section, we first describe the measurement of this band structure evolution under a displacement field.

Measurements were performed in a custom-built dilution refrigerator–based scanning probe system operating at 10 mK (*40*, *41*). Figure 1D shows the STM topograph of the TDBG sample used in this study (fig. S1). The measured moiré periodicity of λ = (8.1 ± 0.1) nm corresponds to a twist angle of θ = (1.74° ± 0.02°), and the local heterostrain was found to be <0.1%, as determined from atomically resolved spatial measurements (*42*). Scanning tunneling spectroscopy (STS) measures the differential tunneling conductance (dI/dV) signal, which is proportional to the local density of states (LDOS) (*43*). For TDBG, the measured dI/dV reflects the LDOS of the topmost layer of the four-layer system. The finite tip potential leads to a fixed top gate V_T, determined by the contact-potential difference between probe and sample. Varying the back-gate voltage V_G simultaneously tunes the displacement field D and the carrier density n following $D \propto C_G V_G - C_T V_T$ and $n \propto C_G V_G + C_T V_T$, where C_G and V_G are the back-gate capacitance and voltage, and C_T and V_T are tip capacitance and potential, respectively (*42*). We mapped the differential tunneling conductance as a function of sample bias V_B, the displacement field D, carrier density n, and magnetic field B. Figure 1E shows the experimental LDOS map at B = 0 T. The low-energy bands V1 and C1 and remote bands V2 and C2 and their

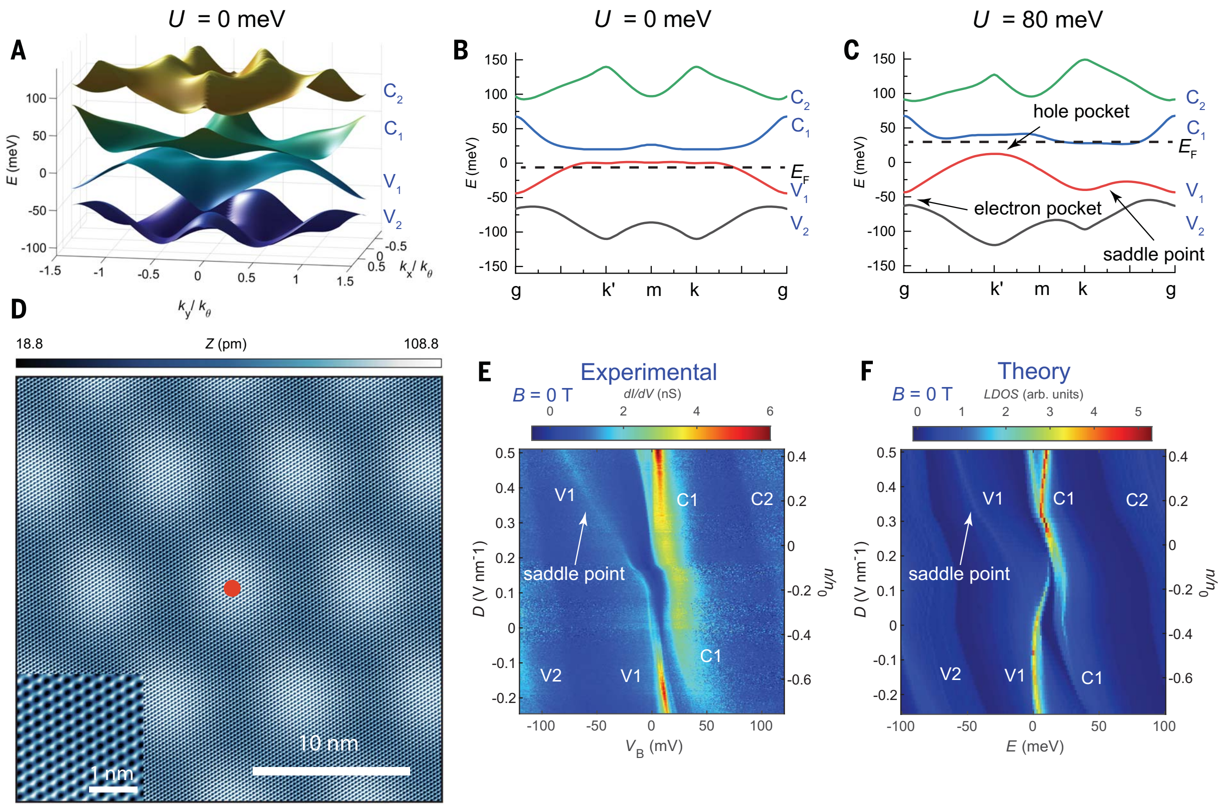


Fig. 1. Tuning the narrow electronic bands in TDBG with displacement field at θ = 1.75°. (**A**) Energy-band structure of TDBG showing the four low-energy bands computed by using a continuum model with interlayer potential U = 0 meV and θ = 1.75°. Line cuts from the energy bands in (A) along Bloch wave vector k_x = 0 for (**B**) U = 0 meV and (**C**) U = 80 meV. A dashed horizontal line indicates the position of the Fermi level (E_F), which moves through the bands as the displacement field and carrier density are varied. The x-axis labels (g, k′, m, and k) denote high-symmetry points in the mini-Brillouin zone. (**D**) STM topographic image of the TDBG moiré pattern, with moiré wavelength λ = (8.1 ± 0.1) nm. (Inset) A magnified portion of the image showing the atomic lattice of graphene. A twist angle of θ = (1.74° ± 0.02°) was determined from the atomically resolved moiré lattice (*42*). The red dot indicates the ABBC symmetry position where the spectral maps with applied magnetic field were acquired. Tunneling setpoint I = 20 pA; V_B = 100 mV; temperature T = 0.01 K. (**E**) Tunneling spectral maps in the D versus V_B plane for B = 0 T. n_0 ($-n_0$) indicates the electron density at full filling of the C1 (V1) band. Setpoints: I = 20 pA, V_B = 120 mV, T = 0.01 K. (**F**) Corresponding theory map for B = 0 T (continuum-model calculation). arb. units, arbitrary units.

evolution with displacement field are clearly visible. The experimental data qualitatively agree with the modeled LDOS of the top layer of the four-layer graphene system obtained with continuum-model calculations [Fig. 1F (*42*)]. The calculations in this work are presented for qualitative comparison and are not meant for quantitative comparison with the experiment.

The trends in the dI/dV map can be qualitatively understood by considering how the theoretically predicted bands in Fig. 1, B and C, fill with n as the Fermi level is raised from negative to positive energies while the bands are modified by the displacement field D. At negative displacement field approaching D = 0, the Fermi level is initially pinned to the flat portions of V1 with its associated high LDOS (Fig. 1B). This appears in the spectral maps (Fig. 1, E and F) with the high LDOS of the V1 band located near the Fermi level, corresponding to zero sample bias. Above a displacement field of 0.2 V/nm, the relatively flat C1 band becomes pinned at the Fermi level, and the V1 band develops a larger dispersion while its highest LDOS at its saddle point moves to lower energies away from C1 with increasing positive displacement field (Fig. 1C). This results in a bright spectral line associated with the saddle point in the V1 band with a negative slope in the top portion of the spectral maps in Fig. 1, E and F.

Mapping Landau levels

Application of a magnetic field creates sharp LLs in the dI/dV spectrum of all four bands, as shown in the dI/dV maps at B = 4 T and 8 T in Fig. 2, B and D, respectively (see figs. S11 to S13 for additional data sets). All dI/dV maps with magnetic field are acquired on an ABBC site (Fig. 1D, red dot). The experimental maps can be compared with the theoretical maps in Fig. 2, C and E, obtained from magnetic field–dependent quantum mechanical calculations (*42*), following the methods in (*44*). A series of LL spectra at different displacement fields are extracted from the experimental map at B = 4 T (Fig. 2A). The LLs in the V1 band are clearly pronounced and show that the LL energies are irregularly spaced, showing neither the equal spacing characteristic for a parabolic band nor the square root dependence on orbital index as observed in graphene. In the next section, we use the evolution of the V1 LL spacing to experimentally demonstrate the characteristics of the tunable band.

The electron-like versus hole-like character of a band is manifested in the energy variation with LL index and magnetic field B. LLs originating from electron-like Fermi pockets increase in energy with LL index and B, whereas the opposite trend applies for hole-like LLs. The saddle-point energy separates electron-like from hole-like pockets. With this in mind, we examined the experimental spectra of Fig. 2A. At negative displacement field, a set of well-resolved V1 LLs is clearly visible in the dI/dV map at B = 4 T (Fig. 2B) and the corresponding extracted spectra (Fig. 2A). We identified the zeroth LL at the bottom of the V1 band, which was at V_B ≈ −45 mV for zero displacement field. For a given displacement field in Fig. 2A, the LL energy spacing in the V1 band is observed to gradually decrease with increasing bias voltage, followed by a gradual increase,

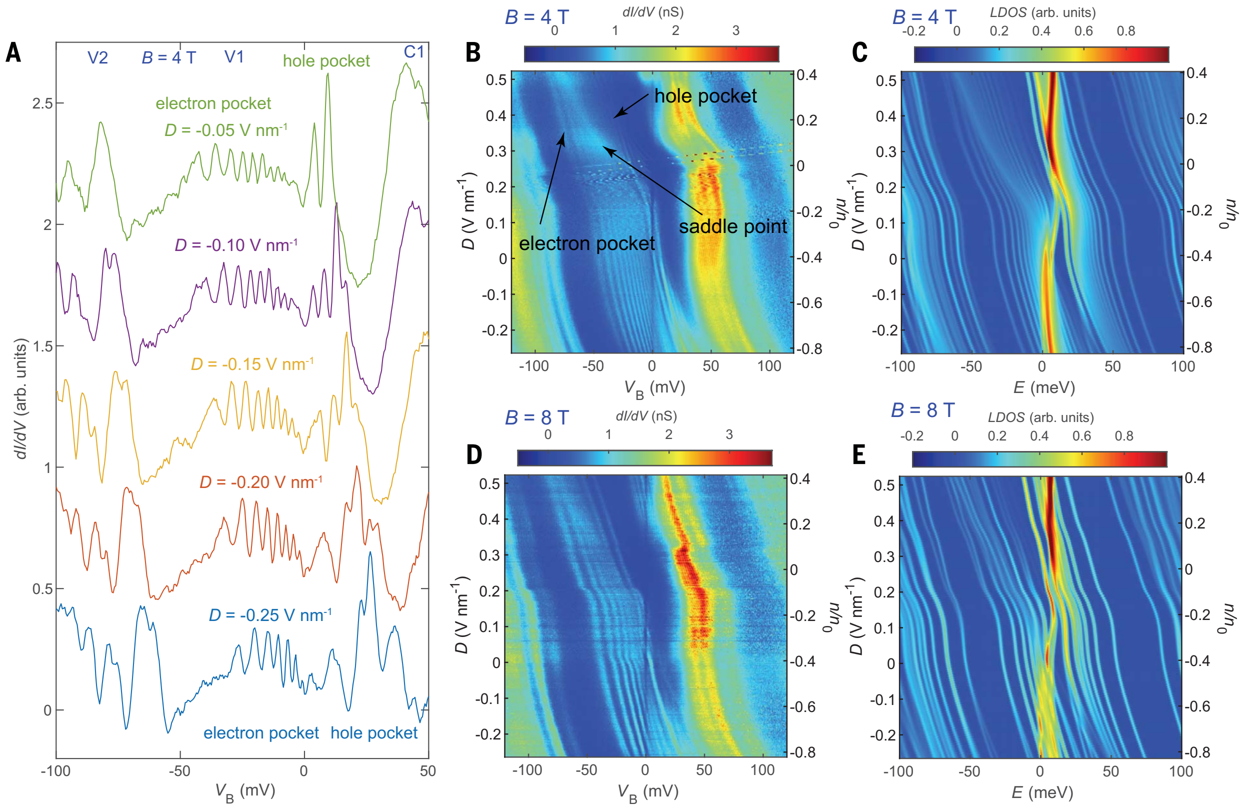


Fig. 2. Tuning LLs in TDBG with displacement and magnetic fields. (**A**) Tunneling spectra versus V_B of TDBG at selected displacement fields taken from the map in (B). Tunneling spectral map in the D versus V_B plane at (**B**) B = 4 T. (**C**) Corresponding theory map for B = 4 T (quantum calculation). (**D**) Tunneling spectral map in the D versus V_B plane at B = 8 T. (**E**) Corresponding theory map for B = 8 T (quantum calculation) (*42*). The maps in (B) and (D) were acquired on the maxima positions in the moiré structure indicated by the red dot in Fig. 1D. Setpoints: I = 20 pA, V_B = 120 mV for (B) and (D), T = 0.01 K. See figs. S11 to S13 for additional spectral map data.

pointing to a transition from an electron-like to a hole-like pocket through a saddle point. The LL spacing is inversely proportional to the LDOS, which peaks at the Van Hove singularity at the saddle point. The saddle point is pinned at the Fermi level for negative carrier densities and visible as the high LDOS crossing through the V1 band at positive densities (Fig. 2, B and D).

We used the B-field dependence of the LLs to confirm the nature of the tunable band structure. Figure 3, A to C, displays dI/dV maps in the B versus V_B planes at a fixed displacement field, showing the dispersion of the LLs with B in the V1 band. At negative displacement fields (Fig. 3, B and C), the LLs below the saddle point at the Fermi level disperse with positive slope, and those above the saddle point disperse with negative slope. This confirms the electron versus hole-pocket nature of the V1 band. The transition from electron to hole pocket in the V1 band is similar at positive displacement fields, where increasing displacement fields tunes the V1 saddle point to lower energies (Fig. 3A). At D = 0.35 V nm^{-1} (Fig. 3A), the saddle point is tuned to the middle of the band with an equal number of positive and negative dispersing LLs on either side, which demonstrates the well-determined tunability of electron and hole pockets with displacement field.

Deviations from the Onsager relation

We now zoom in on the detailed features in the V1 band, leveraging LL spectroscopy to extract quantitative information about the zero-field material properties involving valley-contrasting orbital magnetism and orbital magnetic susceptibility. Figure 3D shows extracted, well-resolved spectra at selected B fields. In Fig. 3E, we show the corresponding peak positions of the LLs in the V1 band versus B for n = 0 to 7, which shows how the LLs disperse with magnetic field. LLs are semiclassically described by Onsager's quantization condition, which requires that the total phase accumulation over a cyclotron orbit be an integer multiple of 2π (*19*). This yields

$$S(E_n)\phi_0/2\pi = 2\pi B_n(n+1/2) \qquad (1)$$

relating the k-space area $S(E_n)$ of the zero-field iso-energy contour of the nth LL E_n to the magnetic field B_n, where $\phi_0 = h/e$ is the magnetic flux quantum, h is Planck's constant, and e is the elementary charge. With the equal-energy k-space area scaling as $B_n(n+1/2)$ for all LLs at energy E_n, a plot of $S \propto [B_n(n+1/2)]_{n=0,1,2...}$ versus LL energy for all LLs would collapse onto a single curve. From Fig. 3E, we plot $S \propto [B_n(n+1/2)]_{n=0...7}$ in Fig. 3F. The fact that experimental data points fail to collapse onto a single curve demonstrates the inadequacy of the Onsager relation and the need for higher-order corrections.

Recently, a systematic expansion of the semiclassical Onsager relation described in Eq. 1 was derived by adding corrections in terms of powers of B (*21*):

$$B_n(n+1/2)/\phi_0 = S(E_n)/4\pi^2 + m'(E_n)B_n + \chi'(E_n)B_n^2/2 + ... \qquad (2)$$

The coefficients of the first- and second-order correction, $m'(E)$ and $\chi'(E)$, are the derivatives with respect to energy of the total orbital magnetic moment and orbital magnetic susceptibility, respectively. Eq. 2 is thus a powerful tool to either infer the zero-field material properties $S(E)$, $m'(E)$, and $\chi'(E)$ from a given LL spectrum, or to predict the detailed LL spectrum given the knowledge of the zero-field material properties. In the next section, we extract these properties for TDBG from the LL spectrum

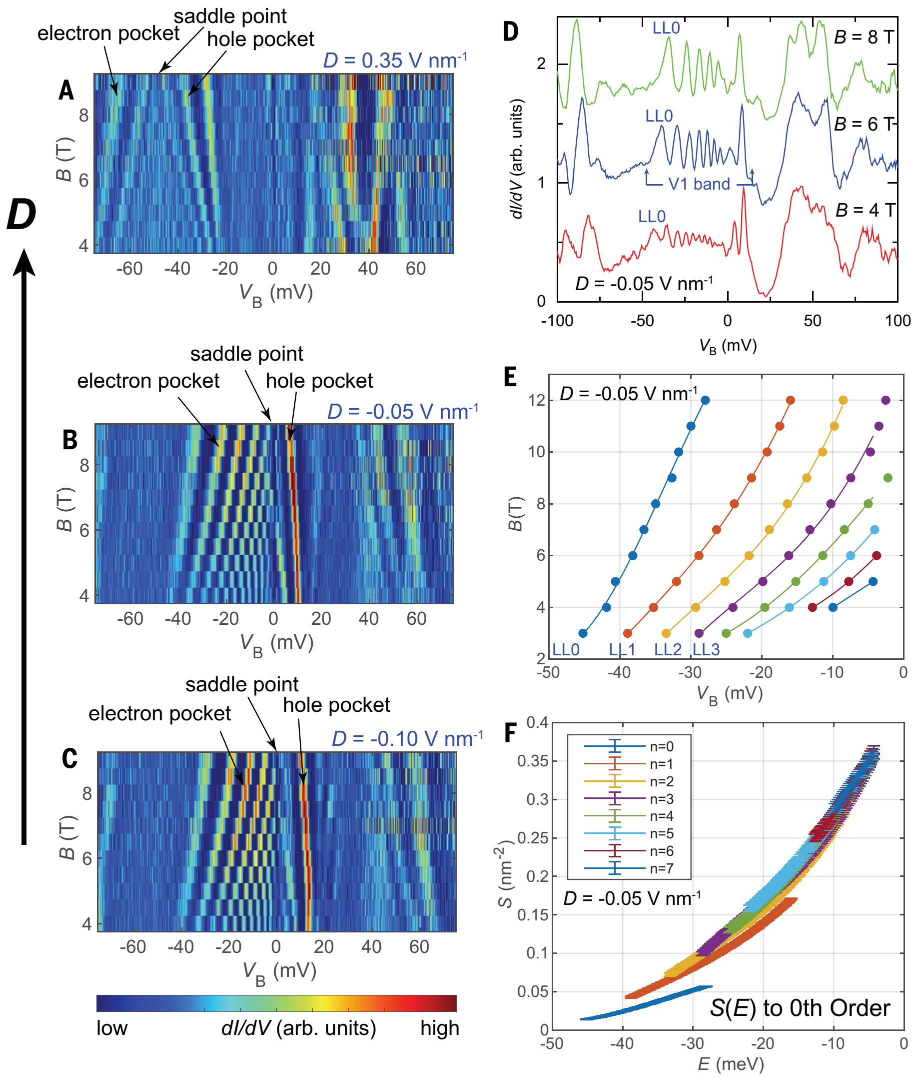


Fig. 3. Breakdown of the standard Onsager relation in TDBG. Spectral maps in the B versus V_B plane for displacement fields of (**A**) D = 0.35 V nm^{-1}, (**B**) D = −0.05 V nm^{-1}, and (**C**) D = −0.1 V nm^{-1} showing the dispersion of the electron and hole-pocket LLs with magnetic field, and the shift of the V1 saddle point to lower energies with increasing displacement field. (**D**) Tunneling spectra versus V_B of TDBG as a function of magnetic field at fixed D = −0.05 V nm^{-1}. (**E**) LL peak positions versus B for D = −0.05 V nm^{-1}. Data from LLs n = 0 to n = 8 are shown. Symbols, experimental peak positions; solid lines, cubic-spline interpolation of experimental peak positions. (**F**) $S(E)$ computed from the solid lines in (E) with Eq. 1. The lack of data collapse indicates that second-order contributions to the LL energies are substantial.

and show a large contribution of $\chi'(E)$ for moiré systems that scale with an increase in moiré lattice constant.

Extracting zero-field properties

The first-order correction $m'(E)$ includes contributions from the total Berry curvature enclosed by an equal energy contour (Fig. 4, A and B), which yields the well-known Berry phase correction, and the average orbital magnetic moment per carrier at energy E (*21*, *45*). The calculated Berry curvature Ω for the V1 band is represented by a hue in the mini-Brillouin zone for each valley K and K' in the Brillouin zone in Fig. 4, A and B, respectively. The valleys have equal and opposite Berry curvature and orbital moment. By symmetry, the total intravalley orbital magnetic moment is finite only at nonzero displacement field. This tunability of the valley-contrasting orbital magnetism enables the disentangling of the orbital susceptibility from the orbital magnetic moment, which we use in the analysis of the LLs below. The strength of these contributions in Eq. 2 can be evaluated by using the continuum model. Figure 4C shows the terms on the right side of Eq. 2 for LL0, calculated with the zero-field continuum model and multiplied by ϕ_0/B to obtain the phase contribution in the extended Onsager relation in Eq. 2 (see fig. S5 for all phase contributions). We note that these phase terms are energy dependent and are large in TDBG; for comparison, a value of 0.5 would be equivalent to the π Berry phase in single-layer graphene. The sign of the orbital moment and its derivative $m'(E)$ is opposite for the K and K' valleys, as expected from the valley-contrasting Berry curvature shown in Fig. 4, A and B.

Figure 4D shows how the phase contributions displayed in Fig. 4C drastically alter the semiclassical LL spectrum. The semiclassical, zeroth-order LL spectrum (Eq. 1) is displayed in blue for the lowest LLs in the V1 band by using $S(E)$ obtained from the zero-field continuum model. The first-order correction $m'(E)$ leads to a valley splitting of the LLs indicated by the dashed red and black lines. The calculated valley splitting at the bottom of the V1 band is substantial but within the LL spacing, where the LL index can be assigned unambiguously. This makes LL0 ideally suitable to determine the experimental valley splitting from our high-resolution measurements.

High-resolution dI/dV spectra of LL0 at increasing magnetic fields 9 to 13 T at D = −0.25 V/nm are shown in Fig. 4E. We observed a clear experimental splitting, with the red lines indicating the fitted peak positions. At 13 T, the LL0 peak is very broad owing to many levels forming from the Hofstadter spectrum at high magnetic fields. We present a comparison to the spectrum computed quantum mechanically in Fig. 4F (gray peaks) to show the onset of Hofstadter minibands at higher fields. The values of splitting obtained with semiclassical calculations are similar (Fig. 4D). Using an energy broadening comparable to the experiment, shown as the black curve, we obtained qualitative agreement with the experiment. Figure 4G shows a direct comparison of the experimental LL0 splitting in Fig. 4E (blue data points) with the quantum mechanical calculations in Fig. 4F (open orange circles) as a function of B field. We found semiquantitative agreement between the calculated and measured valley-splitting values, demonstrating the tunable valley-contrasting orbital magnetism present in TDBG. Most notably, the splitting was much larger than the standard Zeeman effect with Landé g factor of 2 (solid black line) and did not increase linearly with B field as one would expect for a Zeeman-type splitting mechanism. The latter indicates that the splitting is not only given by $m'(B)$, but rather, a considerable nonzero second-order correction χ' leads to nonlinearity in the splitting with B. We obtained an estimate of m' in the middle of the V1 band of ≈ 3.5 × 10^{13}

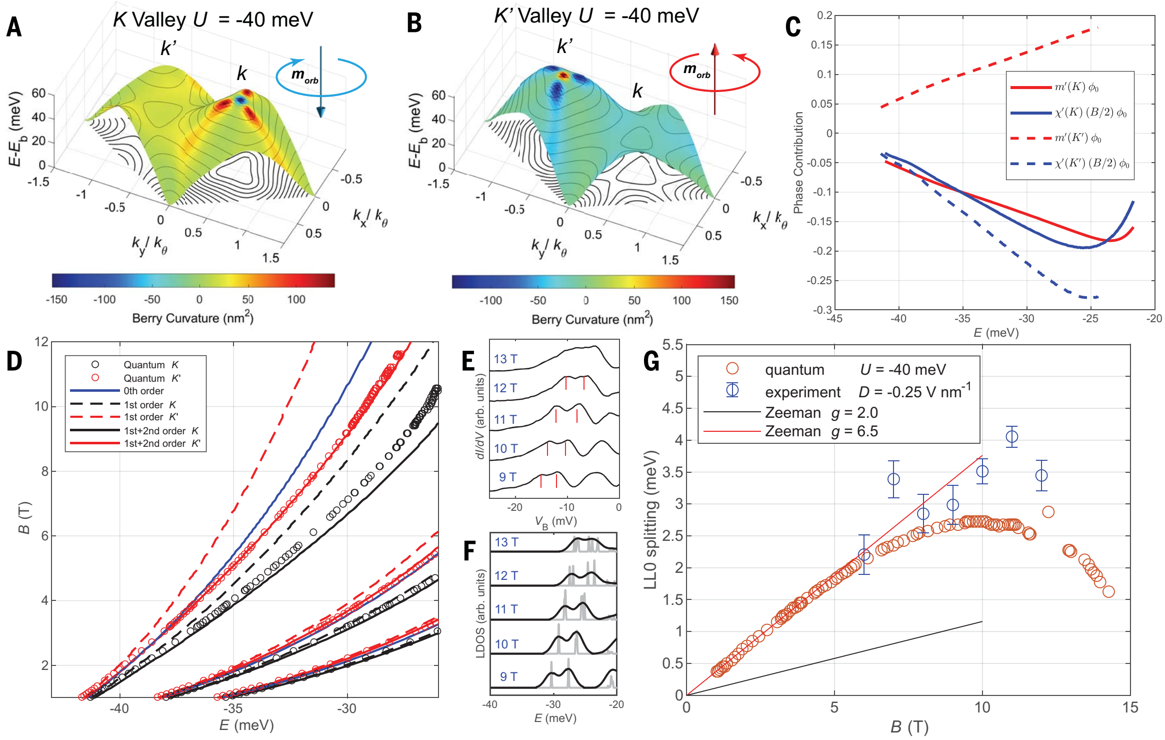


Fig. 4. LL splitting originating from valley-contrasting orbital magnetism and Berry curvature in TDBG. Calculated three-dimensional plot of the V1 band with Berry curvature shading at $U = -40$ meV for (**A**) K and (**B**) K' valleys. K and K' valleys have equal and opposite nonzero Berry curvature and equal and opposite orbital magnetic moment (orbital moments depicted schematically). E_b is the D-dependent energy of the bottom of the V1 band. (**C**) A comparison of the calculated first- (m') and second-order (χ') correction magnitude with the extended Onsager relation for the V1 TDBG band ($U = -40$ meV). See fig. S5 and S6 for additional phase contribution results (*42*). (**D**) Calculation of the first few LLs ($n = 0$, 1, and 2) versus B, comparing quantum mechanical to semiclassical methods with zeroth-, first-, and second-order corrections. The first-order correction is grossly deficient in matching the quantum results, particularly in the $n = 0$ level, whereas including both first- and second-order corrections provides a good agreement with the quantum results ($U = -40$ meV). (**E**) Tunneling spectra of the LL0 peak as a function of B for $D = -0.25$ V nm^{-1} that show a splitting of the LLs in high magnetic fields. The red lines indicate peak positions determined from nonlinear least-square fitting of a double Lorentzian function to LL0. (**F**) Quantum calculation of LL0 as a function of magnetic field for an interlayer potential $U = -40$ meV, corresponding to a displacement field of $D = -0.25$ V nm^{-1}. The LL shows a valley splitting upon application of a displacement field. The solid black lines have a Gaussian energy broadening of $\sigma = 1$ meV to match experiment; gray lines, $\sigma = 0.05$ meV. (**G**) The energy difference of the split $n = 0$ LL from (E) compared with quantum calculations in (F). For the quantum calculation in higher fields, we plot the splitting as the difference in energy between lowest levels in the Hofstadter bands of the K and K' valleys. The uncertainty in the experimental points is derived from twice the standard deviation obtained from the nonlinear least-square fits of the peaks in (E). For comparison, the solid lines show the calculated Zeeman splitting with Landé g factor of (black) $g = 2$ and (red) 6.5.

C/(J s), comparable to a calculated estimate of ≈ 4.4×10^{13} C/(J s), using the values for the splitting, $S'(E)$ and $\chi'(E)$, extracted from the experiment (*42*).

The first-order correction m' does not explain the lack of collapse of the LLs onto a single curve. This is the result of a strong, experimentally observable second-order correction, given by the energy-dependent orbital magnetic susceptibility $\chi'(E)$. This contribution originates from several mechanisms, as elucidated in recent theoretical works (*18*, *46*, *47*), including the geometrical origin, namely the Berry curvature and quantum metric (*48*). The calculated $\chi'(E)$ of TDBG has the same sign for both the K and K' valley, as shown for LL0 in Fig. 4C. This results in a shift of the valley-split LLs, quadratically increasing with B field, as shown by the black and red solid lines in the continuum-model calculation in Fig. 4D. This fully expanded semiclassical approximation with first- and second-order correction (Eq. 2) agrees well with the quantum mechanical spectrum obtained for the same system (red and black open circles), again emphasizing the importance of the second-order correction to adequately describe the system. For a large region of energies and magnetic fields, the second-order correction is larger than the first one (Fig. 4, C and D, and fig. S6). This means that the energy-dependence of the magnetic field–induced orbital moment exceeds the displacement field–induced orbital moment.

To extract the second-order correction to the LL energy, we note that the first-order splitting is typically not visually evident for most electron-like LLs in the V1 band at smaller displacement fields owing to smaller valley splitting from orbital magnetism, and presumably, to broadening. The energy value of a single LL peak is then approximately the average of the valley-split values. The valley-averaged energy at a fixed displacement field is obtained from Eq. 2 by letting $m'(E) = 0$, keeping only the second-order correction. For equal energy LLs n and m with associated fields B_n and B_m, Eq. 2 yields the following for S and χ':

$$S(E) = \frac{2\pi e}{h} B_n B_m \left[\frac{B_m(n+1/2) - B_n(m+1/2)}{(B_m^2 - B_n^2)}\right] \quad (3)$$

$$\chi'(E) = \frac{e}{\pi h}\left[\frac{B_n(n+1/2) - B_m(m+1/2)}{(B_n^2 - B_m^2)}\right] \quad (4)$$

Figure 5A shows the average values of $S(E)$ obtained by evaluating Eq. 3 for all pairs of LLs in Fig. 3E at each energy value. By contrast to Fig. 3F, $S(E)$ collapses onto a single curve (blue data points) in Fig. 5A, confirming the

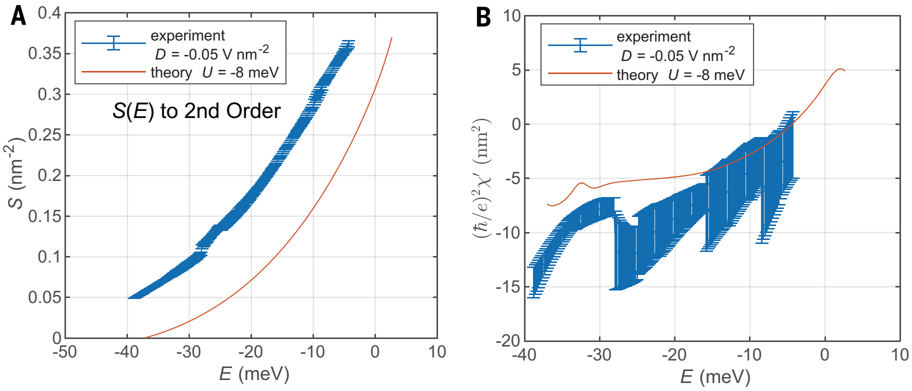


Fig. 5. Extraction of enhanced energy-dependent orbital magnetic susceptibility in moiré quantum matter from the extended Onsager relation. (**A**) The symbols represent $S(E)$ computed from the solid lines in Fig. 3E with Eq. 3, demonstrating data collapse onto one continuous curve. The solid line represents $S(E)$ computed for the V1 band from continuum-model calculations. (**B**) The symbols represent $\chi'(E)$ computed from the solid lines in Fig. 3E with Eq. 4. The solid line represents $\chi'(E)$ computed from continuum-model calculations. See (*42*) for a discussion of the error analysis.

applicability of the second-order correction. The result has a similar trend to that of the zero-field continuum-model calculation of $S(E)$ (orange solid line).

Similarly, we obtained the average value of $\chi'(E)$ by evaluating Eq. 4 (blue data points in Fig. 5B). The experimentally extracted values are in qualitative agreement with the zero-field orbital magnetic susceptibility (orange line), computed with the continuum model according to the formalism of (*49*). The uncertainty in the extracted $\chi'(E)$ and the deviation in overlap of segments from different pairs of LLs are attributed to two factors. First, corrections to Landau quantization energies have a smaller impact on the high-index LLs located at higher energies (*50*). Extracting the relatively smaller corrections at these energies therefore leads to larger uncertainty. Second, variations of the chemical potential with magnetic field shift the LL energy positions, especially at high magnetic fields (*42*). We do not believe interactions make a qualitative impact on our analysis. The LL scaling obtained with the extended Onsager relation is highly constrained, and we find it is well satisfied over a wide range of electron filling. Presumably, interactions would renormalize LL energies in a distinct manner, which would depend strongly on filling factor.

Discussion and outlook

There are several distinct microscopic mechanisms that underlie the orbital magnetic susceptibility, including recently identified geometric contributions (*18, 46*). In fig. S9, we show the contributions to $\chi'(E)$ from these mechanisms [see (*42*) for the full breakdown]. The contributions from the Pauli paramagnetism and the quantum geometry involve the orbital magnetic moment, Berry curvature, and quantum metric. These mechanisms make the dominant contribution to $\chi'(E)$ near the top of V1 and bottom of C1 bands, where the Berry curvature is peaked (Fig. 4, A and B). The contributions from the van Vleck paramagnetism and k-space energy polarization are dominant in the middle of the V1 band, whereas the well-known Landau-Peierls susceptibility is substantial at the band edges.

We attribute the large value of $\chi'(E)$ to the large lattice constant a of the moiré potential. This can be roughly understood with dimensional analysis: for a single band model, the total orbital magnetic susceptibility χ generally scales as Wa^2, where W is the bandwidth, so that the derivative of χ with respect to energy then scales as a^2, which can be large for moiré quantum systems. For a more in-depth analysis, see fig. S10, A and B, which shows that the maximum value of χ' increases dramatically with increasing moiré wavelength, indicating that future measurements of orbital magnetism should be even more observable at larger wavelengths.

MQM holds great opportunities for efficient magnetization control, topological transport phenomena, the quantum anomalous Hall effect, and technological applications in magnetoelectric, magneto-optics, and topological spintronics. Understanding the underlying band topology and magnetic response functions are the key metrological capability required to practically harness all these opportunities. Our measurements and analysis point a way forward to this goal.

REFERENCES AND NOTES

1. National Academies of Sciences, Engineering, and Medicine, *Theoretical Considerations Concerning Moiré Quantum Materials in Frontiers in Synthetic Moiré Quantum Matter: Proceedings of a Workshop* (2022); https://nap.nationalacademies.org/catalog/26594/frontiers-in-synthetic-moire-quantum-matter-proceedings-of-a-workshop.
2. D. M. Kennes *et al.*, *Nat. Phys.* **17**, 155–163 (2021).
3. R. Bistritzer, A. H. MacDonald, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 12233–12237 (2011).
4. Y. Cao *et al.*, *Nature* **556**, 80–84 (2018).
5. Y. Cao *et al.*, *Nature* **556**, 43–50 (2018).
6. J. M. Park, Y. Cao, K. Watanabe, T. Taniguchi, P. Jarillo-Herrero, *Nature* **590**, 249–255 (2021).
7. Y. Cao, J. M. Park, K. Watanabe, T. Taniguchi, P. Jarillo-Herrero, *Nature* **595**, 526–531 (2021).
8. H. Zhou, T. Xie, T. Taniguchi, K. Watanabe, A. F. Young, *Nature* **598**, 434–438 (2021).
9. J. M. Park *et al.*, *Nat. Mater.* **21**, 877–883 (2022).
10. G. W. Burg *et al.*, *Nat. Mater.* **21**, 884–889 (2022).
11. Y. Zhang *et al.*, *Science* **377**, 1538–1543 (2022).
12. Y.-H. Zhang, D. Mao, Y. Cao, P. Jarillo-Herrero, T. Senthil, *Phys. Rev. B* **99**, 075127 (2019).
13. M. V. Berry, *Proc.- Royal Soc., Math. Phys. Eng. Sci.* **392**, 45–57 (1984).
14. A. L. Sharpe *et al.*, *Science* **365**, 605–608 (2019).
15. M. Serlin *et al.*, *Science* **367**, 900–903 (2020).
16. S. Sinha *et al.*, *Nat. Commun.* **11**, 5548 (2020).
17. D. Xiao, M.-C. Chang, Q. Niu, *Rev. Mod. Phys.* **82**, 1959–2007 (2010).
18. Y. Gao, S. A. Yang, Q. Niu, *Phys. Rev. B Condens. Matter Mater. Phys.* **91**, 214405 (2015).
19. L. Onsager, *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **43**, 1006–1008 (1952).
20. L. M. Roth, *Phys. Rev.* **145**, 434–448 (1966).
21. Y. Gao, Q. Niu, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7295–7300 (2017).
22. C. Shen *et al.*, *Nat. Phys.* **16**, 520–525 (2020).
23. X. Liu *et al.*, *Nature* **583**, 221–225 (2020).
24. M. He *et al.*, *Nat. Phys.* **17**, 26–30 (2021).
25. Y. Cao *et al.*, *Nature* **583**, 215–220 (2020).
26. P. C. Adak *et al.*, *Phys. Rev. B* **101**, 125428 (2020).
27. S. Sinha *et al.*, *Nat. Phys.* **18**, 765–770 (2022).
28. C. Zhang *et al.*, *Nat. Commun.* **12**, 2516 (2021).
29. X. Liu *et al.*, *Nat. Commun.* **12**, 2732 (2021).
30. C. Rubio-Verdú *et al.*, *Nat. Phys.* **18**, 196–202 (2021).
31. P. Rickhaus *et al.*, *Science* **373**, 1257–1260 (2021).
32. J. Y. Lee *et al.*, *Nat. Commun.* **10**, 5333 (2019).
33. M. Koshino, *Phys. Rev. B* **99**, 235406 (2019).
34. N. R. Chebrolu, B. L. Chittari, J. Jung, *Phys. Rev. B* **99**, 235417 (2019).
35. Y. W. Choi, H. J. Choi, *Phys. Rev. B* **100**, 201402 (2019).
36. F. J. Culchac, R. R. Del Grande, R. B. Capaz, L. Chico, E. S. Morell, *Nanoscale* **12**, 5014–5020 (2020).
37. F. Haddadi, Q. Wu, A. J. Kruchkov, O. V. Yazyev, *Nano Lett.* **20**, 2410–2415 (2020).
38. J. A. Crosse, N. Nakatsuji, M. Koshino, P. Moon, *Phys. Rev. B* **102**, 035421 (2020).
39. M. Kuiri *et al.*, *Nat. Commun.* **13**, 6468 (2022).
40. Y. J. Song *et al.*, *Rev. Sci. Instrum.* **81**, 121101 (2010).
41. J. Schwenk *et al.*, *Rev. Sci. Instrum.* **91**, 071101 (2020).
42. Additional supplementary text and data are available in the supplementary materials.
43. J. A. Stroscio, W. J. Kaiser, *Scanning Tunneling Microscopy: Volume 27 (Methods of Experimental Physics)* (Academic Press, 1993).
44. R. Bistritzer, A. H. MacDonald, *Phys. Rev. B Condens. Matter Mater. Phys.* **84**, 035440 (2011).
45. M.-C. Chang, Q. Niu, *J. Phys. Condens. Matter* **20**, 193202 (2008).
46. F. Piéchon, A. Raoux, J.-N. Fuchs, G. Montambaux, *Phys. Rev. B* **94**, 134423 (2016).
47. M. Ogata, H. Fukuyama, *J. Phys. Soc. Jpn.* **84**, 124708 (2015).
48. F. Wilczek, A. Shapere, *Geometric Phases in Physics* in *Advanced Series in Mathematical Physics* (World Scientific, 1989), vol. 5; https://www.worldscientific.com/worldscibooks/10.1142/0613.
49. H. Fukuyama, *Prog. Theor. Phys.* **45**, 704–729 (1971).
50. J.-N. Fuchs, F. Piéchon, G. Montambaux, *SciPost Physics* **4**, 024 (2018).
51. M. R. Slot *et al.*, Replication Data for A Quantum Ruler for Orbital Magnetism in Moiré Quantum Matter, George Mason University Dataverse (2023); https://doi.org/10.13021/orc2020/RLZTDF.

ACKNOWLEDGMENTS

We thank F. J. Giessibl for providing the qPlus AFM sensors and general advice on AFM measurements. **Funding:** University of Maryland and the National Institute of Standards and Technology (NIST) PREP program grant no. 70NANB18H165 (Y.M., E.M.S., and E.S.); Georgetown University and NIST PREP program grant no. 70NANB18H161 (M.R.S.); University of Maryland and NIST Joint Quantum Institute grant no. 70NANB21H126 (S.K. and D.Y.); Office of Naval Research grant no. N00014-20-1-2352 (S.K. and D.Y.); Japan Society for the Promotion of Science KAKENHI grant nos. 19H05790, 20H00354, and 21H05233 (T.T. and K.W.); Dutch Research Council (NWO) through Rubicon grant no. 019.193EN.026 (M.R.S.); University of Reno start-up grant no. PG19012 (Y.B.). **Author contributions:** Conceptualization: F.G., P.M.H., J.A.S., and N.B.Z.; Methodology: P.M.H. and Y.B.; Investigation: M.R.S., Y.M., S.K., D.T.W., E.S., S.T.L., E.M.S., D.Y., S.R.B., and F.G.; Visualization: Y.M., M.R.S., P.M.H., and J.A.S.; Project administration: J.A.S.; Supervision: J.A.S. and N.B.Z.; Resources: K.W. and T.T.; Software: P.M.H., Y.B., M.R.S., Y.M., S.K., D.T.W., and J.A.S.; Writing – original draft: M.R.S., Y.M., P.M.H., N.B.Z., F.G., and J.A.S.; Writing – review and editing: M.R.S., Y.M., P.M.H., S.K., D.T.W., E.S., S.T.L., E.M.S., D.Y., S.R.B., K.W., T.T., T.B., N.B.Z., F.G., and J.A.S. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The data from this study and the code used to generate theoretical results in this study are available at the George Mason University Dataverse (*51*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf2040
Materials and Methods
Supplementary Text
Figs. S1 to S13
References (*52–56*)

Submitted 5 October 2022; resubmitted 1 March 2023
Accepted 30 August 2023
10.1126/science.adf2040

QUANTUM INFORMATION

An atomic-scale multi-qubit platform

Yu Wang[1,2]†, Yi Chen[1,2,3,4]†, Hong T. Bui[1,5]†, Christoph Wolf[1,2], Masahiro Haze[1,6], Cristina Mier[1,7], Jinkyung Kim[1,5], Deung-Jang Choi[1,7,8,9], Christopher P. Lutz[10], Yujeong Bae[1,5]*, Soo-hyon Phark[1,2]*, Andreas J. Heinrich[1,5]*

Individual electron spins in solids are promising candidates for quantum science and technology, where bottom-up assembly of a quantum device with atomically precise couplings has long been envisioned. Here, we realized atom-by-atom construction, coherent operations, and readout of coupled electron-spin qubits using a scanning tunneling microscope. To enable the coherent control of "remote" qubits that are outside of the tunnel junction, we complemented each electron spin with a local magnetic field gradient from a nearby single-atom magnet. Readout was achieved by using a sensor qubit in the tunnel junction and implementing pulsed double electron spin resonance. Fast single-, two-, and three-qubit operations were thereby demonstrated in an all-electrical fashion. Our angstrom-scale qubit platform may enable quantum functionalities using electron spin arrays built atom by atom on a surface.

Constructing and coherently controlling nanoscale qubit systems lies at the very heart of quantum-coherent nanoscience (*1*). This length scale requires the use of fundamental quantum properties of atoms, such as the spin of electrons, which naturally occurs in many solid-state environments and allows high-fidelity operations and readout by electromagnetic means (*2*). Despite decades of effort, however, it remains a formidable task to realize an atomic-scale quantum architecture in which multiple electron spin qubits can be precisely assembled, controllably coupled, and coherently operated. Electron spin qubits created in dopants in semiconductors and color centers in insulators, for example, can be well controlled individually (*3*, *4*) but are difficult to couple together into a circuit. An attractive alternative approach is to use atomic-level fabrication with a scanning tunneling microscope (STM), in which atom manipulation (*5*) or selective desorption (*6*) can lead to designed quantum spin architectures (*7*, *8*). As a first step toward in situ operation of atomic quantum devices, a radiofrequency (RF) voltage was used to coherently control a single electron spin in the STM tunnel junction (*9*, *10*) in a so-called electron spin resonance (ESR)–STM setup (*11*–*14*). However, achieving the potential of this platform requires multiple addressable and detectable qubits that lie outside of the subnanometer tunnel junction region.

Building a multi-qubit platform atom by atom

Figure 1A illustrates our strategy to construct such an atomic-scale multi-qubit platform. A "sensor" qubit consisting of a spin-1/2 hydrogenated Ti atom on a bilayer MgO film (*14*, *15*) was positioned in the tunnel junction, driven, and detected by a magnetic tip (*16*, *17*). Outside of the tunnel junction, addressable "remote" qubits were created by positioning a spin-1/2 Ti atom (*14*, *15*) ~0.6 nm away from a single Fe atom (*18*) using atom manipulation [see section 1 of (*19*)]. An Fe atom can be seen as a single-atom magnet in this work because its spin relaxation time far exceeds the time scale of qubit operations and its resonance frequencies are far off the applied RF range (*11*, *18*, *20*) [see section 2 of (*19*)]. The main purpose of Fe atoms here is to substitute for the tip in providing a static local magnetic field for remote Ti spins (those not directly under the tip apex) to allow their electrical driving. Under a tip-supplied RF electric field, Ti spins are believed

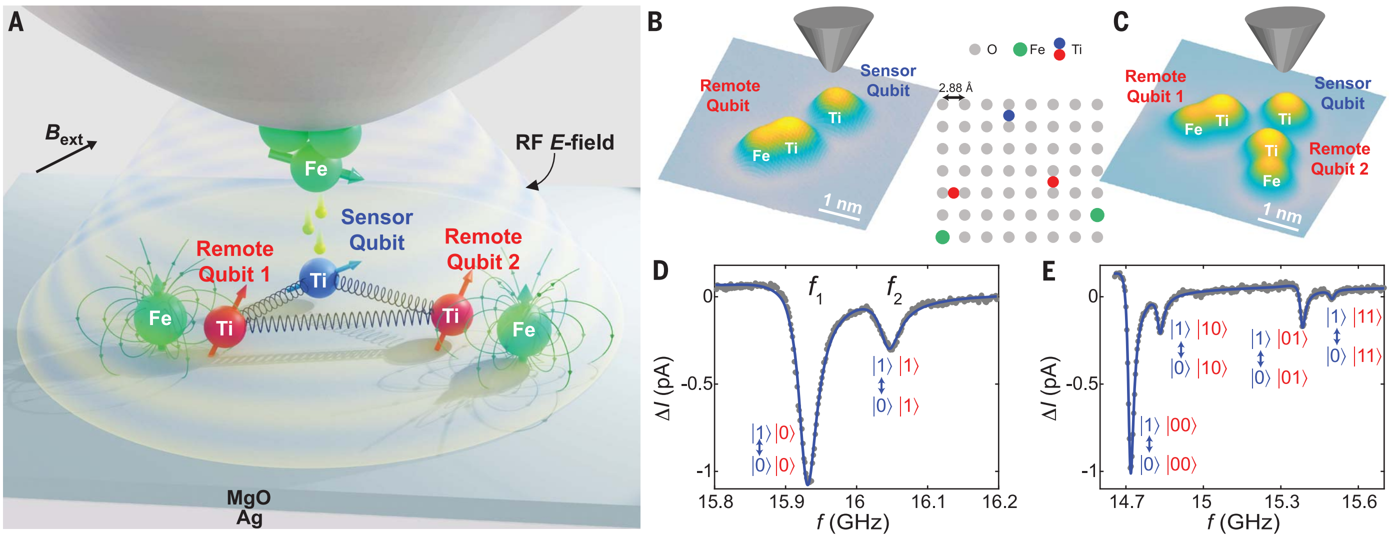


Fig. 1. Bottom-up construction of multiple coupled electron spin qubits. (**A**) Schematic: A sensor spin qubit (Ti, blue) is placed under the apex of a spin-polarized STM tip for readout. Remote qubits are constructed at precise separations to the sensor qubit by atom manipulation. Each remote qubit is composed of a spin-1/2 Ti atom (red) and a single-atom magnet (Fe) (green), where Fe's magnetic field gradient, in combination with the RF electric field between the tip and the sample, coherently drives remote qubits. (**B** and **C**) Constant-current STM images showing atom-by-atom construction of a multi-qubit structure composed of two (B) and three (C) qubits (image size: 5.0 × 5.0 nm). Inset: atomic registry of the structure in (C). Structure in (B) has the same configuration but without remote qubit 2. (**D** and **E**) Continuous-wave ESR spectra measured with the tip positioned on the sensor qubit in the two-qubit (D) and three-qubit (E) structure. The quantum states of the sensor qubit are labeled by the first kets. Each ESR transition of the sensor qubit distinguishes a quantum state of the remote qubits (the second kets). The measured spin-polarized tunnel-current signal ΔI is the difference between averaged signals in lock-in A and B subcycles (fig. S10), which reflects the change of the sensor spin polarization due to applied coherent control pulses [see section 4 of (*19*)]. Imaging conditions in (B) and (C): sample bias voltage V_{DC} = 100 mV, time-averaged tunnel current I_{DC} = 10 pA. ESR conditions in (D) and (E): V_{DC} = 50 mV, I_{DC} = 20 pA, zero-to-peak RF voltage V_{RF} = 30 mV. The sample was kept at 0.4 K during measurements.

to undergo relative oscillations inside Fe's magnetic field gradient, thus experiencing an effective resonant driving magnetic field (*21*, *22*), in analogy to micromagnet driving in quantum dots (*23*) [more details of the driving mechanism can be found in section 2 of (*19*) and in (*24*)]. By contrast, ESR of the sensor qubit, as well as previous ESR-STM spectroscopy (*12*–*14*), relies on the tip's magnetic field gradient (*21*, *22*). Initialization of the qubits was performed thermally by cooling the sample to 0.4 K and applying an external magnetic field, producing a thermal spin state having a predominant population in the spin ground state (~90% under a 0.7 T field).

A representative structure composed of two remote qubits and one sensor qubit was constructed atom by atom using atom manipulation [see section 1 of (*19*)], as shown in Fig. 1, B and C. The qubit-qubit couplings, including dipolar and exchange spin-spin couplings, are sensitive to their atomic separations down to the angstrom scale (*14*, *15*) (figs. S1 and S2), consistent with results in other material systems (*25*). The STM-based atomically precise construction scheme thus allowed us to engineer the resonant frequencies and couplings among all of the spins, an essential step for addressing and detecting multiple qubits individually [see section 3 of (*19*) and fig. S2].

Detection of the remote qubits was achieved through ESR spectroscopy of the sensor qubit, the ESR transition frequency of which depends on the quantum states of other qubits (Fig. 1, D and E). The ESR frequencies of the qubits were designed to be sufficiently separated compared with the qubit-qubit couplings so that the multi-qubit states could be well described by Zeeman product states, for which we used the first (second) brackets to denote the quantum states of the sensor (remote) qubit. For example, in a structure composed of a sensor qubit and a remote qubit (Fig. 1D), the two ESR transitions of the sensor qubit at frequencies f_1 (corresponding to transition $|0\rangle|0\rangle \leftrightarrow |1\rangle|0\rangle$) and f_2 ($|0\rangle|1\rangle \leftrightarrow |1\rangle|1\rangle$) detect the populations of $|0\rangle$ and $|1\rangle$ states of the remote qubit, respectively. The difference between f_1 and f_2 is determined by the Ti-Ti coupling [see section 3 of (*19*)]. Multiple remote qubits can be simultaneously sensed in a similar fashion (Fig. 1E).

Single-qubit operation and readout of a remote qubit

To demonstrate this qubit control and readout scheme, we began by measuring a remote qubit's ESR spectrum with the tip positioned above the sensor qubit, as sketched in Fig. 2A. The energy levels and ESR transitions of this two-qubit structure are illustrated in Fig. 2B, and its STM image is shown in Fig. 1B. To individually address the sensor and remote qubits, we used two RF sources to apply two consecutive RF voltage pulses to the STM tip.

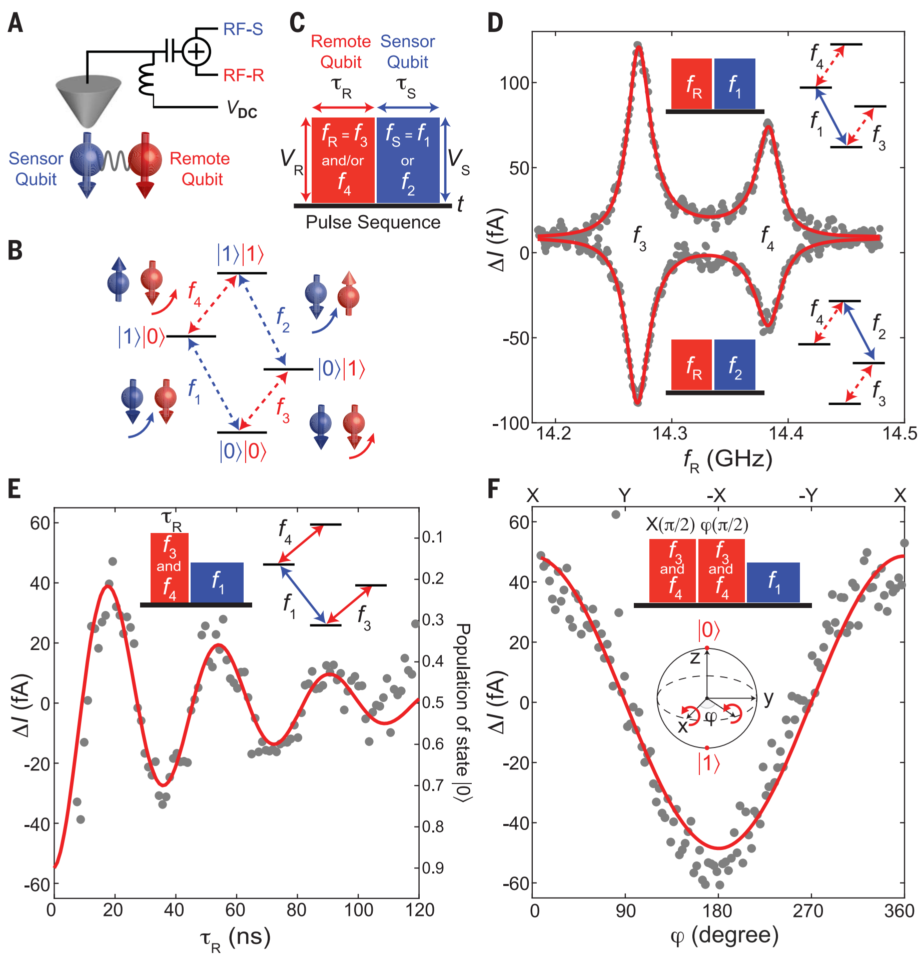


Fig. 2. Coherent control of a single remote qubit. (**A**) Schematic of the measurement scheme. Two RF tones, RF-S and RF-R, were applied to the STM tip for the coherent control of the sensor and remote qubit, respectively. The atomic structure used in this figure is shown in Fig. 1B. (**B**) Energy diagram and ESR transitions of the two-qubit system. (**C**) Typical pulse sequence composed of a control pulse on the remote qubit (red) followed by a sensing pulse on the sensor qubit (blue). The control pulse can induce conditional or unconditional qubit rotations depending on its frequency content. (**D**) ESR spectra of the remote qubit measured with the tip positioned on the sensor qubit. The two curves correspond to different sensing frequencies (top, $f_S = f_1$; bottom, $f_S = f_2$). Insets: pulse sequence and ESR transitions involved in each spectrum. (**E**) Rabi oscillations of the remote qubit performed by simultaneously driving $f_R = f_3$ and f_4 to induce unconditional qubit rotations. The sensing pulse was subsequently applied at $f_S = f_1$. Red curve is a fit to an exponentially decaying sinusoid. A second *y* axis on the right shows the population of state $|0\rangle$, which is determined by temperature and transition frequency at $\tau_R = 0$ (here 90%) (*3*). (**F**) Two-axis control of the remote qubit shown by sweeping the relative phase ϕ between two consecutive $\pi/2$ pulses ($f_R = f_3$ and f_4; $f_S = f_1$). The data have been vertically shifted by 300 fA to allow easier comparison to (E). ϕ = 180 degrees does not correspond to $\Delta I = 0$ here because of the RF rectification effect [see section 4 of (*19*) and fig. S10]. Solid curve shows a cosine fit. The axes of the two-step qubit rotation are illustrated by red arrows on the Bloch sphere in the rotating frame. ESR conditions in (D): V_{DC} = 50 mV, I_{DC} = 20 pA, $V_S = V_R$ = 30 mV, $\tau_R = \tau_S$ = 200 ns; and in (E) and (F): V_{DC} = 50 mV, I_{DC} = 20 pA, V_S = 50 mV, V_R = 120 mV, τ_S = 200 ns. The sample was kept at 0.4 K during measurements.

[1]Center for Quantum Nanoscience, Institute for Basic Science (IBS), Seoul 03760, Korea. [2]Ewha Womans University, Seoul 03760, Korea. [3]International Center for Quantum Materials, School of Physics, Peking University, Beijing 100871, China. [4]Collaborative Innovation Center of Quantum Matter, Beijing 100871, China. [5]Department of Physics, Ewha Womans University, Seoul 03760, Korea. [6]The Institute for Solid State Physics, University of Tokyo, Kashiwa 277-8581, Japan. [7]Centro de Física de Materiales CFM/MPC (CSIC-UPV/EHU), 20018 Donostia-San Sebastián, Spain. [8]Donostia International Physics Center (DIPC), 20018 Donostia-San Sebastián, Spain. [9]Ikerbasque, Basque Foundation for Science, 48013 Bilbao, Spain. [10]IBM Research Division, Almaden Research Center, San Jose, CA 95120, USA.
***Corresponding author. Email: bae.yujeong@qns.science (Y.B.); phark@qns.science (S.P.); heinrich.andreas@qns.science (A.J.H.)**
†These authors contributed equally to this work.

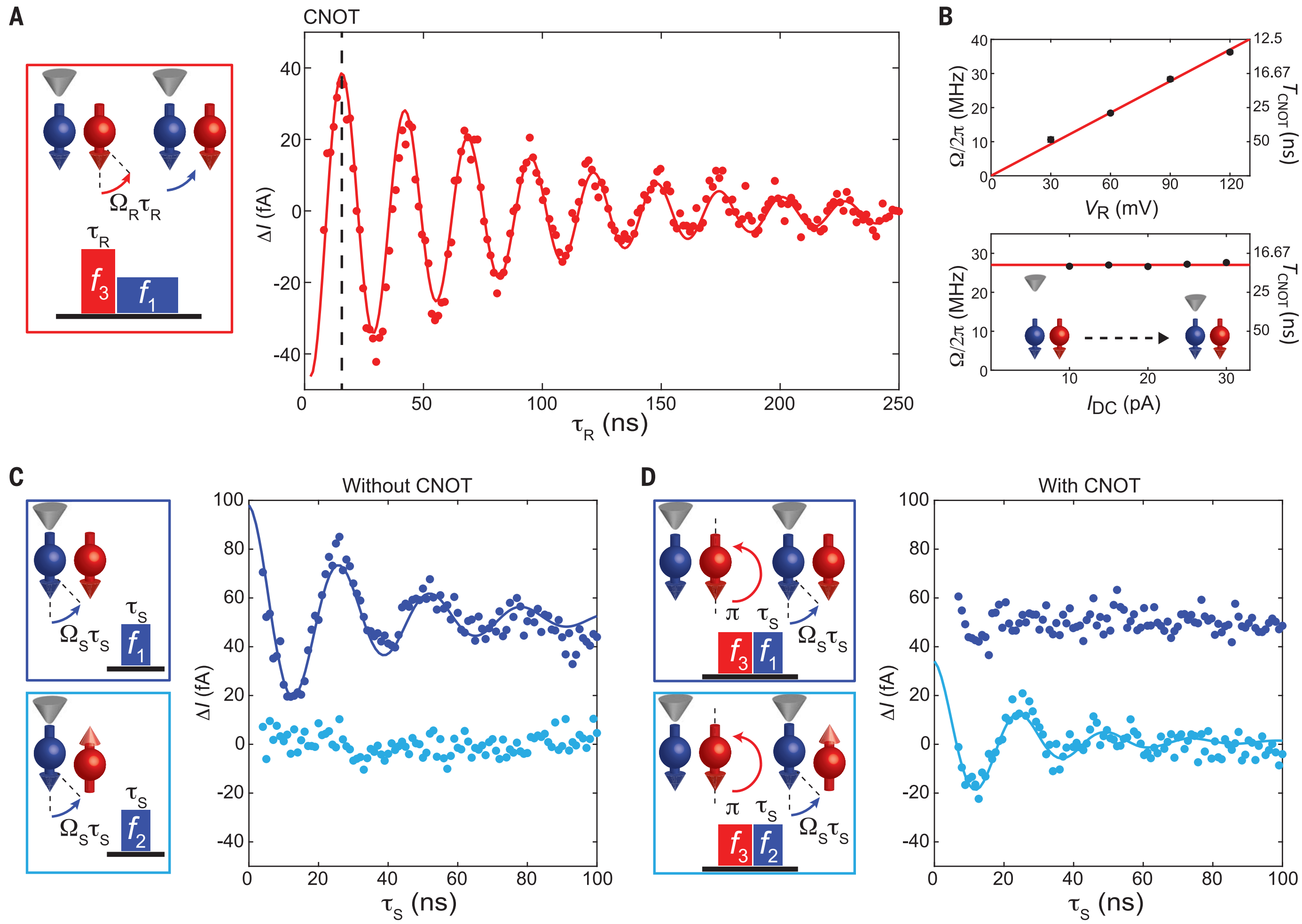


Fig. 3. Two-qubit operations in the coupled-qubit structure shown in Fig. 1B. (**A**) Controlled rotation of the remote qubit, conditional on the sensor qubit state being |0⟩, obtained by driving the transition |0⟩|0⟩ ↔ |0⟩|1⟩ at $f_R = f_3$. Red points show coherent oscillations of the remote qubit detected through a sensing pulse at $f_S = f_1$, as shown schematically on the left. Red curve is a fit to an exponentially decaying sinusoid. The CNOT operation time is ~13 ns (dotted line). (**B**) Rate of controlled rotations, Ω/2π, of the remote qubit. The rate increased linearly with the RF voltage V_R (upper) but remained unchanged as the tunnel current I_{DC} (lower) was varied. Error bars are comparable with the size of symbols. Corresponding CNOT times are plotted on a second *y* axis on the right. Solid curves are linear fits. Inset illustrates the change of tip-sensor separations. (**C** and **D**) Controlled rotations of the sensor qubit without (C) and with (D) a CNOT operation at $f_R = f_3$ on the remote qubit. Blue (cyan) points measured with $f_S = f_1$ (f_2) show sensor Rabi oscillations (solid lines are exponentially decaying sinusoidal fits) only when the remote qubit has a significant population in state |0⟩ (|1⟩). The CNOT operation transferred the predominant population from state |0⟩|0⟩ to state |0⟩|1⟩, thus causing the opposite trends of the oscillations between (C) and (D). Blue curves are shifted vertically by 50 fA for clarity. ESR conditions in (A): V_{DC} = 50 mV, I_{DC} = 10 pA, V_S = 30mV, V_R = 120 mV, τ_S = 200 ns; in (C): V_{DC} = 20 mV, I_{DC} = 7.5 pA, V_S = 100 mV; and in (D): V_{DC} = 20 mV, I_{DC} = 7.5 pA, V_S = V_R = 100 mV. The sample was kept at 0.4 K during measurements.

A control pulse was applied at frequency f_R to control the remote qubit, followed by a sensing pulse at frequency f_S acting on the sensor qubit (Fig. 2, B and C). A sensing pulse applied at $f_S = f_1$ (f_2) results in an ESR signal that depends on the population of state |0⟩ (|1⟩) of the remote qubit. To obtain the ESR spectrum of the remote qubit, we swept the frequency f_R of the control pulse across the remote qubit resonances while keeping the sensing pulse fixed at a resonance frequency of the sensor qubit. When f_R matched an ESR transition of the remote qubit (e.g., |0⟩|0⟩ ↔ |0⟩|1⟩), the joint two-qubit state populations were altered, resulting in a detectable increase (decrease) of the sensor's ESR signal at f_1 (f_2) [see section 4 of (*19*) and Fig. 2D]. This measurement is conceptually similar to ensemble double-electron spin resonance spectroscopy and allowed us to directly obtain the resonance frequencies of the remote qubit (f_3 and f_4; Fig. 2D) even though no tunnel current passed through it.

Single-qubit control can be performed in our platform by coherently driving a qubit independent of other qubits' quantum states. This was achieved by exciting all ESR transitions of a remote qubit with the same driving strength (*26*), as illustrated in the insets of Fig. 2, E and F. We measured Rabi oscillations of a remote qubit by varying the pulse duration τ_R of the control pulse and subsequently sensing at f_1 (Fig. 2E).

Two-axis control (*27*) of the remote qubit can be demonstrated by varying the relative phase ϕ of two consecutive π/2 pulses (Fig. 2F). On the Bloch sphere in the rotating frame, the first π/2 pulse rotates the qubit state from the *z* axis to the *y* axis, and the second performs a π/2 rotation around an axis in the *x*–*y* plane at an angle of ϕ to the *x* axis (Fig. 2F, bottom inset). The resulting signal was well described by the expected cos(ϕ) dependence (*27*). Two-axis control allows arbitrary single-qubit operations (*28*).

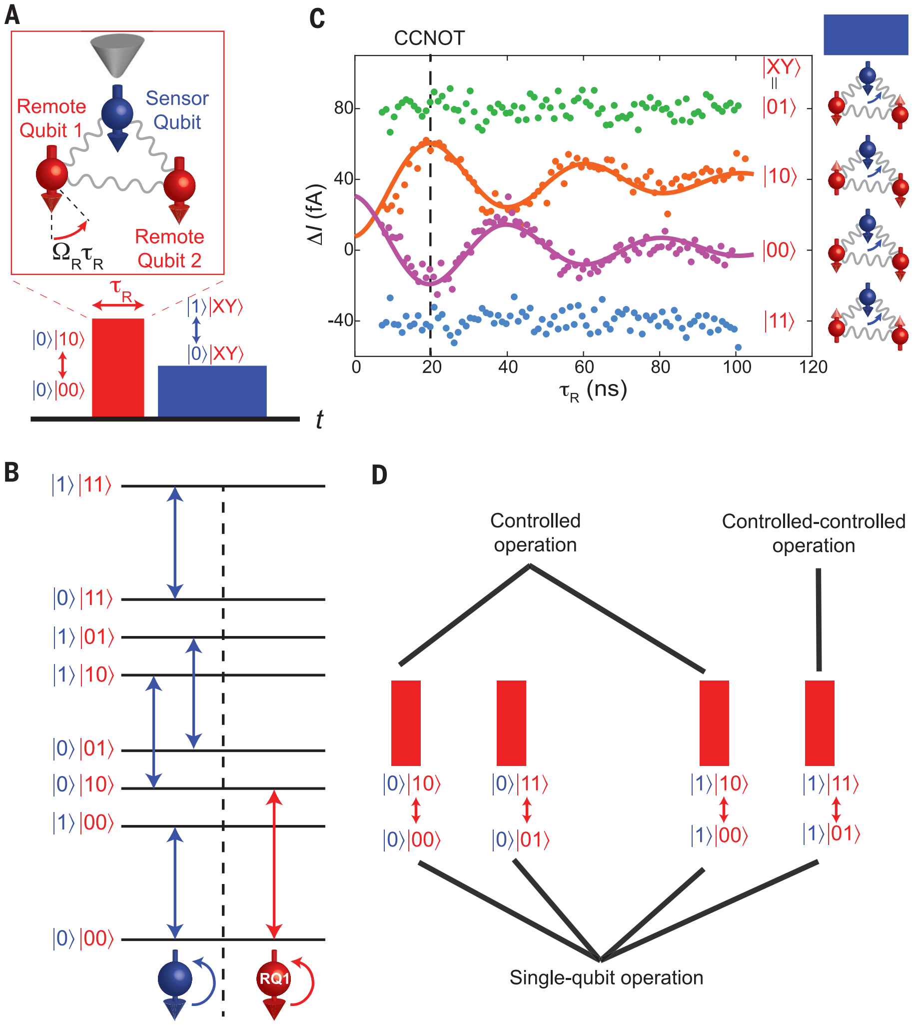


Fig. 4. Three-qubit operations in a multi-qubit atomic structure. (**A**) Schematic of the control scheme of a multi-qubit structure (Fig. 1C) composed of two remote qubits and a sensor qubit. (**B**) Energy diagram and ESR transitions used for the measurement in (C). (**C**) Controlled-controlled operation of remote qubit 1 performed by driving the transition |0⟩|00⟩ ↔ |0⟩|10⟩ [(A), red pulse; (B), red arrow]. Four different frequencies of the sensor qubit were used for sensing the four remote qubit states [(A), blue pulse; (B), blue arrows, where the detected states of the remote qubits are labeled by red kets in (C)]. The two oscillating curves in (C) (orange and magenta) correspond to detection of the remote qubits' |00⟩ or |10⟩ states upon the controlled-controlled operation at |0⟩|00⟩ ↔ |0⟩|10⟩. The CCNOT operation time is ~20 ns (dotted line). The other two qubit states were not driven so they showed no oscillations. Green, orange, and blue curves are vertically shifted for clarity. (**D**) Strategy to perform single-, two-, and three-qubit operations in our platform. Single-qubit operations are performed by nonselectively exciting all ESR transitions of a qubit with the same strength. Two-qubit controlled operations were performed by selectively exciting ESR transitions that correspond to a specific quantum state of one control qubit. Three-qubit controlled-controlled operations are performed by selectively exciting ESR transitions that correspond to a specific quantum state of multiple control qubits. ESR conditions in (C): V_{DC} = 50 mV, I_{DC} = 20 pA, V_S = 50 mV, V_R = 80 mV, τ_S = 200 ns. The sample was kept at 0.4 K during measurements.

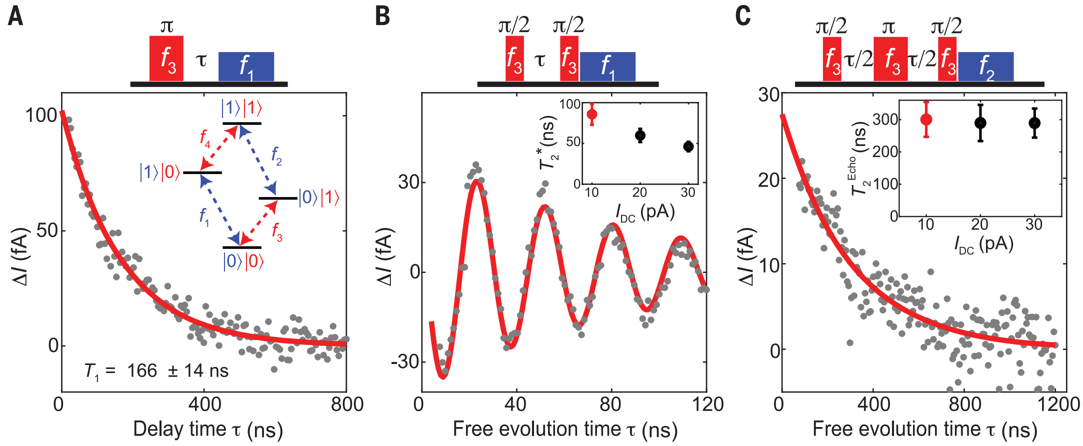


Fig. 5. Relaxation and coherence properties of remote qubits. The two-qubit structure shown in Fig. 1B was used in these measurements. (**A**) Relaxation time T_1 of the remote qubit measured by an inversion recovery scheme. The exponential fit (solid line) yields a relaxation time T_1 of 166 ± 14 ns for the remote qubit. Inset: Labels of ESR transitions used in this figure (same as Fig. 2B). (**B**) Ramsey measurements of the remote qubit with $f_S = f_1$ and $f_R = f_3$ + 30 MHz yielding a coherence time T_2^* = 86 ± 13 ns. Inset: T_2^* shows a dependence on the tunnel current I_{DC}. (**C**) Spin-echo measurements of the remote qubit measured with $f_S = f_2$ and $f_R = f_3$. The exponential fit (solid line) yields a coherence time of T_2^{Echo} = 300 ± 54 ns. Inset: T_2^{Echo} shows no dependence on I_{DC}. ESR conditions in (A): V_{DC} = 50 mV, I_{DC} = 20 pA, V_S = 50 mV, V_R = 120 mV, τ_S = 200 ns; and in (B) and (C): V_{DC} = 50 mV, I_{DC} = 10 pA, V_S = 60 mV, V_R = 120 mV, τ_S = 200 ns. The sample was kept at 0.4 K during measurements.

Two-qubit controlled operation

Controlled NOT (CNOT) operations are straightforward to perform in our platform by selectively exciting ESR transitions that correspond to a specific quantum state of a control qubit. In the case of a two-qubit structure, we could perform a CNOT operation of the remote qubit by driving a single ESR transition such as |0⟩|0⟩ ↔ |0⟩|1⟩ at $f_R = f_3$ (Fig. 3A), which changes the state of the remote qubit if and only if the sensor qubit is in state |0⟩. Here, the quantum state of the remote qubit was subsequently

measured by applying a sensing pulse at f_1 (Fig. 3A). Sensing at f_2 (instead of at f_1) showed oscillations of the opposite sign because this effectively detected the population of state $|1\rangle$ (instead of state $|0\rangle$) of the remote qubit (fig. S4) (*26*). The oscillations in Fig. 3A show a CNOT operation time of ~13 ns for the remote qubit. The gate operation rate increased in proportion to the RF amplitude (Fig. 3B, top, and fig. S3A), with an upper limit set by RF heating effects. The fast operation of remote qubits is believed to result from the strong magnetic field gradient of the neighboring Fe atom (*24*). As a consequence of Fe-dominated driving, the CNOT rate is independent of how far the tip is to the remote qubits, as reflected by the tunnel current I_{DC} (Fig. 3B, bottom, and fig. S3B) (*24*).

To evaluate the effect of the CNOT operation on the remote qubit, we performed controlled rotations of the sensor qubit without and with this CNOT operation (Fig. 3, C and D, respectively). Because the initial thermal population of the two-qubit system was predominantly in state $|0\rangle|0\rangle$, a controlled rotation of the sensor qubit at f_1 starting from the initial state $|0\rangle|0\rangle$ (i.e., without a CNOT operation) showed clear oscillations because a pulse at f_1 excites the transition $|0\rangle|0\rangle \leftrightarrow |1\rangle|0\rangle$. By contrast, measurement at transition f_2 showed no detectable oscillations (Fig. 3C) because remote qubit state $|1\rangle$ was nearly unoccupied. These coherent oscillations were reversed after a CNOT operation of the remote qubit at f_3 (Fig. 3D), which transferred the predominant population from state $|0\rangle|0\rangle$ to state $|0\rangle|1\rangle$, hence becoming accessible to a sensing pulse at $f_S = f_2$ (i.e., $|0\rangle|1\rangle \leftrightarrow |1\rangle|1\rangle$) but not f_1. We further verified that a CNOT operation of the remote qubit at f_4 (instead of f_3) did not strongly affect the oscillations observed on the sensor qubit (fig. S6), highlighting the selective, controlled nature of CNOT operations.

Multi-qubit controlled operation

Although single- and two-qubit operations are sufficient to generate arbitrary quantum circuits, multi-qubit operations can significantly reduce the execution time and mitigate accumulated operation errors (*29*). As long as the qubit-qubit couplings can be spectroscopically resolved, our platform allows fast, native qubit operations having multiple control qubits (by selectively exciting certain ESR transitions, as shown in Fig. 4D). To demonstrate this ability, we constructed a three-qubit structure composed of two remote qubits (referred to as RQ1 and RQ2; see fig. S2 for details) and a sensor qubit (as sketched in Fig. 4A and imaged in Fig. 1C). We selectively drove RQ1's $|0\rangle|00\rangle \leftrightarrow |0\rangle|10\rangle$ transition (Fig. 4B, red arrow), corresponding to a rotation of RQ1 if and only if the sensor qubit and RQ2 were both in state $|0\rangle$ (Fig. 4A). When varying the pulse duration τ_R, this controlled-controlled operation caused oscillating populations between states $|00\rangle$ and $|10\rangle$ of the two remote qubits, whereas populations of states $|01\rangle$ and $|11\rangle$ were left unchanged (Fig. 4C). Here, the changes of populations were determined from the intensities of the four ESR transitions of the sensor qubit (Fig. 4A, blue pulse, and Fig. 4B, blue arrows). Similarly, the controlled-controlled operation $|0\rangle|01\rangle \leftrightarrow |0\rangle|11\rangle$ caused oscillating populations between states $|01\rangle$ and $|11\rangle$ of the two remote qubits, albeit with reduced amplitudes because of the reduced initial thermal populations in these states (fig. S7E). From these measurements, we obtained a controlled-controlled NOT (CCNOT) operation time as short as 20 ns. Single-, two-, and three-qubit gates in our platform have comparably fast operation times because they only differ in the frequency content of the RF pulses (*30*) (Fig. 4D).

Characterizing the coherence of remote qubits

In STM-based approaches, tunneling electrons have posed severe limitations on the energy relaxation time T_1 and coherence time T_2 of the spins (*18*, *31*). This limitation is overcome in our scheme because tunneling electrons pass only through the sensor qubit and do not flow through the remote qubits (Fig. 1A). To characterize the remote qubits, we implemented an inversion recovery measurement by first applying a π pulse on the remote qubit to invert its population. After a delay time τ, we applied a sensing pulse to the sensor qubit to measure the state of the remote qubit and obtained an energy relaxation time $T_1 = 166 \pm 14$ ns (*14*). A pronounced improvement can be seen in the quantum coherence of remote qubits, which is already visible in the higher quality of the remote qubit's Rabi oscillations (Fig. 3A) compared with the sensor qubit (Fig. 3, C and D). Figure 5B shows the Ramsey signal measured at a tunnel current of $I_{DC} = 10$ pA, from which we extracted a coherence time $T_2^* = 86 \pm 13$ ns. We found that T_2^*, the coherence time subject to inhomogeneous broadening, depended on the tunnel current and thus on the tip height (Fig. 5B, inset, and fig. S8A), suggesting that the proximity of the STM tip to the remote qubits influenced their quantum coherence time despite the absence of tunnel current through them. This decoherence effect was likely caused by slow thermal or field fluctuations, the latter possibly arising from slight, uncontrolled tip motions (*31*). To cancel the effect of this inhomogeneous broadening, we performed a spin-echo measurement and observed that the measured T_2^{Echo} showed negligible dependence on the tunnel current (Fig. 5C, inset, and fig. S8B). The measured coherence time $T_2^{\text{Echo}} = 300 \pm 54$ ns (Fig. 5C) approaches the theoretical limit of $2T_1$. These measurements highlight that the quantum coherence of remote qubits after spin-echo filtering is limited by energy relaxation events, in contrast to other solid-state qubits (*3*, *27*), where $T_1 \gg T_2^{\text{Echo}}$ usually applies.

Discussion and outlook

Our work introduces an atom-by-atom constructed qubit platform using electron spins on surfaces where cryogenic initialization, universal multi-qubit operations, and multi-qubit detection have been demonstrated. Two types of future developments will improve the performance of this qubit platform. The first seeks to obtain more gate operations within the spin decoherence time. Our fast gate operation of ~20 ns implies that a significant improvement can be attained with a moderate increase of T_2, e.g., through the use of thicker insulators (*18*), a detection mechanism that avoids magnetic tips (*32*), the depletion of surrounding nuclear spins (~5% of atoms in MgO have nuclear spins), and the use of quantum gates that are protected by dynamical decoupling-schemes (*33*). For reference, electron spins in bulk insulators (*34*) and semiconductors (*35*) have been demonstrated with coherence times of ~1 ms.

The second development seeks to increase the number of controllable qubits. Given that the RF electric field from the tip can have a radiation range of tens of nanometers (corresponding to the effective tip diameter) (*36*), scaling up the number of remote qubits should not be limited by driving but rather by detection. A direct extension of the work shown here would lead to up to four remote qubits being placed around a sensor qubit for direct read-out. Further increase will rely on remote qubits at next-nearest neighbor sites, where a SWAP gate (*28*) could be used to transfer the quantum information (*37*). Finally, the computational Hilbert space can be further expanded by using molecular qubits to construct three-dimensional architectures (*38*) or by using qudit systems with multiple electron and nuclear spin states (*39*, *40*). Such extensions rely on the specific benefit of our surface-based approach, in which a myriad of available spin species (*17*, *41*–*43*) and geometries (*44*, *45*) can be precisely assembled using scanning probe techniques.

REFERENCES AND NOTES

1. A. J. Heinrich *et al*., *Nat. Nanotechnol.* **16**, 1318–1329 (2021).
2. D. D. Awschalom, R. Hanson, J. Wrachtrup, B. B. Zhou, *Nat. Photonics* **12**, 516–527 (2018).
3. M. T. Mądzik *et al*., *Nat. Commun.* **12**, 181 (2021).
4. P. Neumann *et al*., *Nat. Phys.* **6**, 249–253 (2010).
5. D. M. Eigler, E. K. Schweizer, *Nature* **344**, 524–526 (1990).
6. M. Kiczynski *et al*., *Nature* **606**, 694–699 (2022).
7. H. J. Lee, W. Ho, M. Persson, *Phys. Rev. Lett.* **92**, 186802 (2004).
8. E. Liebhaber *et al*., *Nat. Commun.* **13**, 2160 (2022).
9. K. Yang *et al*., *Science* **366**, 509–512 (2019).
10. P. Willke *et al*., *ACS Nano* **15**, 17959–17965 (2021).
11. S. Baumann *et al*., *Science* **350**, 417–420 (2015).
12. T. S. Seifert *et al*., *Sci. Adv.* **6**, eabc5511 (2020).
13. M. Steinbrecher *et al*., *Phys. Rev. B* **103**, 155405 (2021).
14. L. M. Veldman *et al*., *Science* **372**, 964–968 (2021).
15. Y. Bae *et al*., *Sci. Adv.* **4**, eaau4159 (2018).
16. R. Wiesendanger, *Rev. Mod. Phys.* **81**, 1495–1550 (2009).
17. Y. Chen, Y. Bae, A. J. Heinrich, *Adv. Mater.* **35**, e2107534 (2023).

18. W. Paul *et al*., *Nat. Phys.* **13**, 403–407 (2016).
19. See the supplementary materials.
20. P. Willke *et al*., *Nano Lett.* **19**, 8201–8206 (2019).
21. J. L. Lado, A. Ferrón, J. Fernández-Rossier, *Phys. Rev. B* **96**, 205420 (2017).
22. K. Yang *et al*., *Phys. Rev. Lett.* **122**, 227203 (2019).
23. Y. Tokura, W. G. van der Wiel, T. Obata, S. Tarucha, *Phys. Rev. Lett.* **96**, 047202 (2006).
24. S.-H. Phark *et al*., *Adv. Scie*2302033 (2023).
25. B. Koiller, X. Hu, S. Das Sarma, *Phys. Rev. Lett.* **88**, 027903 (2002).
26. W. Huang *et al*., *Nature* **569**, 532–536 (2019).
27. E. Kawakami *et al*., *Nat. Nanotechnol.* **9**, 666–670 (2014).
28. G. Wolfowicz, J. J. L. Morton, *eMagRes* **5**, 1515–1528 (2016).
29. M. D. Reed *et al*., *Nature* **482**, 382–385 (2012).
30. L. M. Vandersypen *et al*., *Nature* **414**, 883–887 (2001).
31. P. Willke *et al*., *Sci. Adv.* **4**, eaaq1543 (2018).
32. L. Sellies, R. Spachtholz, P. Scheuerer, J. Repp, Atomic-scale access to long spin coherence in single molecules with scanning force microscopy. arXiv:2212.12244 [cond-mat.mes-hall] (2022).
33. D. Suter, G. A. Álvarez, *Rev. Mod. Phys.* **88**, 041001 (2016).
34. E. D. Herbschleb *et al*., *Nat. Commun.* **10**, 3766 (2019).
35. J. T. Muhonen *et al*., *Nat. Nanotechnol.* **9**, 986–991 (2014).
36. Z. Chu, L. Zheng, K. Lai, *Annu. Rev. Mater. Res.* **50**, 105–130 (2020).
37. Y. P. Kandel *et al*., *Nature* **573**, 553–557 (2019).
38. X. Zhang *et al*., *Nat. Chem.* **14**, 59–65 (2022).
39. K. Yang *et al*., *Nat. Nanotechnol.* **13**, 1120–1125 (2018).
40. S. Thiele *et al*., *Science* **344**, 1135–1138 (2014).
41. G. Czap *et al*., *Science* **364**, 670–673 (2019).
42. S. Kovarik *et al*., *Nano Lett.* **22**, 4176–4181 (2022).
43. R. Kawaguchi *et al*., *Nano Lett.* **23**, 213–219 (2023).
44. A. A. Khajetoorians, D. Wegner, A. F. Otte, I. Swart, *Nat. Rev. Phys.* **1**, 703–715 (2019).
45. K. Yang *et al*., *Nat. Commun.* **12**, 993 (2021).
46. Y. Wang, Data and codes for: An atomic-scale multi-qubit platform, Zenodo (2023); https://doi.org/10.5281/zenodo.7984789.

ACKNOWLEDGMENTS

We thank A. Ardavan, M. Ternes, J. F. Rossier, D. Loss, F. Donati, and F. Cho for fruitful discussions. **Funding:** This work was supported by the Institute for Basic Science (grant IBS-R027-D1). C.P.L. acknowledges support from the Office of Naval Research (grant N00014-21-1-2467). M.H. acknowledges support from the University of Tokyo Global Activity Support Program for Young Researchers (FY2020). D.-J.C. and C.M. acknowledge support from the Spanish State Research Agency (grants RTI2018-097895-B-C44, Excelencia EUR2020-112116, and PID2021-127917NB-I00) funded by MCIN/AEI/10.13039/501100011033 and by "ERDF: A way of making Europe" and Eusko Jaurlaritza (project PIBA 2020_1_0017). Y.C. acknowledges support from the National Natural Science Foundation of China (grant 12250001). **Author contributions:** A.J.H., S.P., and C.P.L. conceived the project. Y.W., Y.C., H.T.B., M.H., C.M., J.K., D.-J.C., Y.B., and S.P. performed the measurements and analyzed the data. C.W. performed simulations to optimize the measurement schemes. All authors discussed and prepared the manuscript together. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data that support the plots within the paper and other findings of this study are present in the main and supplementary figures. Raw data and Matlab codes are available from Zenodo (*46*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade5050
Materials and Methods
Supplementary Text
Figs. S1 to S10
References (*47*, *48*)

Submitted 20 August 2022; resubmitted 4 March 2023
Accepted 30 August 2023
10.1126/science.ade5050

PHOTOSYNTHESIS

A chlorophyll *c* synthase widely co-opted by phytoplankton

Yanyou Jiang[1]†, Tianjun Cao[1,2]†, Yuqing Yang[1], Huan Zhang[1,2], Jingyu Zhang[1], Xiaobo Li[1,2]*

Marine and terrestrial photosynthesis exhibit a schism in the accessory chlorophyll (Chl) that complements the function of Chl *a*: Chl *b* for green plants versus Chl *c* for most eukaryotic phytoplankton. The enzymes that mediate Chl *c* biosynthesis have long remained elusive. In this work, we identified the CHLC dioxygenase (Phatr3_J43737) from the marine diatom *Phaeodactylum tricornutum* as the Chl *c* synthase. The *chlc* mutants lacked Chl *c*, instead accumulating its precursors, and exhibited growth defects. In vitro, recombinant CHLC protein converted these precursors into Chl *c*, thereby confirming its identity. Phylogenetic evidence demonstrates conserved use of CHLC across phyla but also the existence of distinct Chl *c* synthases in different algal groups. Our study addresses a long-outstanding question with implications for both contemporary and ancient marine photosynthesis.

Photosynthesis has shaped and continues to shape our landscape, oceans, and atmosphere. Photosynthetic creatures from terrestrial and aquatic environments differ in their light-harvesting systems. In land plants and certain cyanobacteria, chlorophyll (Chl) *b* is used to complement the absorption spectrum of the commonly harbored Chl *a*; by contrast, many eukaryotic algal groups, despite being morphologically and phylogenetically diverse, have consistently adopted Chl *c* as the accessory chlorophyll (*1*). The existence of Chl *c* as the third chlorophyll species was reported as early as 1864 (*2*). However, its biosynthetic mechanism was poorly understood, in contrast to that for cyanobacterial Chl *f*, whose synthase was identified shortly after its discovery (*3*, *4*).

Phylogenetically, Chl *c*–containing algae together with some of their nonphotosynthetic relatives are termed the CASH lineages (where CASH stands for cryptophytes, alveolates, stramenopiles, and haptophytes) (*5*). Among them, cryptophytes notably harbor four genome compartments; many dinoflagellates (alveolates) cause toxic red tides or serve as endosymbionts in corals; diatoms (stramenopiles) contribute 20% of global primary productivity; brown algae (also stramenopiles) make up kelp forests as habitats of marine creatures; and coccolithophores (haptophytes) precipitate carbon dioxide from the air into rocks (*6*, *7*) (fig. S1A). Thus, Chl *c* is of great relevance in contemporary marine photosynthesis and global biogeochemistry. However, our poor understanding of Chl *c* biosynthesis and, as a result, the absence of mutants with a primary defect in Chl *c* accumulation have hindered our direct characterization of the functions of Chl *c*. Furthermore, synthetic biology efforts to transfer the blue-green light–capturing, Chl *c*–containing antennae (*6*) to other phototrophs is not possible without knowing the identity of the Chl *c* synthase.

Elucidation of Chl *c* biosynthesis is also highly relevant to our understanding of the evolution of aquatic phototrophs. All Chl *c*–containing algae harbor plastids of red algal (eukaryotic) origins; yet, whether they acquired the plastid from a single red alga engulfment event or serially has been debated (*8*) (fig. S1, B and C). The identification of the enzymes involved in the biosynthesis of Chl *c*—a pigment shared by these algae but lacking in red algae—will shed light on the evolutionary history of these red plastid–bearing algae (*9*). Recent advancements in the diatom genetic toolbox offer opportunities for studying Chl *c* biosynthesis.

A predicted oxidoreductase, CHLC, is required for Chl c_2 and Chl c_1 accumulation in a diatom

We aimed to identify the Chl *c* synthase from the genetically tractable marine diatom *Phaeodactylum tricornutum*, which harbors Chl c_2 and Chl c_1 (*10*). We first applied CRISPR-Cas9 technologies that were recently developed for this alga (*11*) to obtain mutants for candidate genes. On the basis of structural features, researchers have hypothesized that Chl c_2 could be derived from the common chlorophyll precursor divinyl protochlorophyllide *a* (DVP) (*12*) through a single desaturation reaction, whereas Chl c_1 could be generated through the same conversion on monovinyl protochlorophyllide *a* (MVP) or through reduction of the C8 vinyl group in Chl c_2 by divinyl reductases (*1*, *13*) (Fig. 1). Thus, we screened for genes containing oxidoreductase domains whose mutation would abolish the accumulation of the two Chl *c* species and prioritized those coexpressed with known chlorophyll biosynthesis genes or light-harvesting complex (*LHC*) genes (*14*) (data S1 and materials and methods).

For 20 genes, we successfully obtained mutant transformants, but none of them bore a visible color phenotype (data S1). By contrast, for the gene annotated to encode a phytanoyl-CoA hydroxylase (PHYH; Phatr3_J43737; renamed CHLC, as explained below) (Fig. 2A), many transformants appeared greenish (Fig. 2B), which contrasts with the brown color of wild-type (WT) *P. tricornutum* cells. PHYH (EC 1.14.11.18) was originally discovered in mammals (*15*), and it is a 2-oxoglutarate (2OG)–dependent dioxygenase that catalyzes hydroxylation of a carbon atom. However, enzymes of this family can catalyze a variety of reactions. For example, a bacterial 2OG-dependent dioxygenase, the carbapenem synthase CarC (EC 1.14.20.3), catalyzes the formation of a C=C double bond at the α- and β-positions of the carboxyl group (*16*), similar to the scenario of Chl *c* generation from DVP or MVP (Fig. 1). We thus focused on characterizing the potential function of CHLC in Chl *c* biosynthesis.

To ascertain that the green coloration resulted from the disruption of *CHLC*, we genotyped both green- and brown-colored transformants. We found that all green-colored mutants harbored frameshift-causing mutations or large deletions in the *CHLC* gene (fig. S2; primers in data S2), and we subcloned three randomly chosen colonies to yield the pure, stably maintained *chlc-1*, *chlc-2*, and *chlc-3* mutant lines (Fig. 2, D, E, and G). By contrast, all the brown transformants showed the WT sequence, indicating that their respective *CHLC* gene sequences were not disrupted (fig. S2). Complementation of a representative mutant, *chlc-1*, with a copy of *CHLC* on an episome that encodes a WT CHLC protein fused with the green fluorescent protein (GFP) at its C terminus restored the WT appearance (Fig. 2, C to G), which demonstrates that CHLC is required for the normal coloration of *P. tricornutum* cells.

To identify the specific metabolites affected by *CHLC* mutations, we performed pigment analyses using high-performance liquid chromatography (HPLC) (data S3). We observed that *chlc* mutants exhibited abolished Chl c_2 and c_1 accumulation and a lowered fucoxanthin–to–Chl *a* ratio (Fig. 2H). These results may reflect a combination of primary and secondary effects of CHLC loss, in which the primary outcome of CHLC loss is on one of the light-harvesting antenna pigments (Chl *c* or fucoxanthin) and the secondary impact is on the accumulation of the other one, as we have recently reported (*17*).

Our HPLC pigment analyses also revealed an accumulation of peaks that we assigned as DVP and MVP on the basis of their absorbance profiles (*18*) (Fig. 2H). Because of the similar absorbance spectra of Chl c_2, Chl c_1, DVP, and MVP as well as the isomeric relationship between Chl c_1 and DVP, we further analyzed their fragmentation patterns in liquid chromatography–

[1]Research Center for Industries of the Future, Key Laboratory of Growth Regulation and Translational Research of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, China. [2]Institute of Biology, Westlake Institute for Advanced Study, Hangzhou, China.
*Corresponding author. Email: lixiaobo@westlake.edu.cn
†These authors contributed equally to this work.

tandem mass spectrometry (LC–MS/MS) (fig. S3). The two pigments assigned as Chl c_2 and Chl c_1 were purified together from WT cells and analyzed as a single sample. They both exhibited the $[M+H–18]^+$ dehydration fragment in their secondary mass spectra; this fragmentation was only reported in chlorophylls with an acrylic side chain at C17 (*19*). By contrast, the pigments assigned as DVP and MVP were purified from the *chlc-1* mutant, and neither of them gave rise to $[M+H–18]^+$ fragment peaks under the same conditions (fig. S3). These results confirm our identification of the four compounds.

The accumulation of DVP and MVP but not fucoxanthin precursors is opposite to our observations in fucoxanthin biosynthesis mutants (*17*). Thus, these results suggest that CHLC is directly involved in Chl *c*—but not fucoxanthin—biosynthesis.

Recombinant CHLC produces Chl c_2 and c_1 from precursors accumulated in *chlc* mutants

To directly test the function of CHLC in Chl *c* biosynthesis, we investigated its activity in vitro. We produced recombinant CHLC proteins (labeled PtCHLCs) and incubated them with the DVP-MVP mixture isolated from the *chlc-1* mutant. Chl c_2 and c_1 were readily produced (Fig. 3A). The conversion depended on the presence of 2OG and ferrous iron and was hampered when ascorbate was omitted. These requirements are consistent with other 2OG-dependent dioxygenases (*16*), which suggests that the CHLC reaction involves hydrogen atom transfer followed by electron transfer, as observed for the 2OG-dependent desaturase CarC (*20*, *21*). Next, we performed an additional round of purification to obtain DVP free of MVP to be used as the substrate and found that CHLC was able to convert DVP into Chl c_2 (Fig. 3B); this cleaner assay further confirms the assignment of CHLC desaturase activity illustrated in Fig. 1. Therefore, on the basis of the above results, we propose to rename this protein Chl *c* synthase (CHLC).

We next explored the mechanistic basis of certain structural features of Chl *c* (*22*). Unlike Chls *a*, *b*, *d*, and *f*, which contain a chlorin skeleton (with a reduced D ring), the two major Chl *c* species, Chl c_2 and Chl c_1, retain the porphyrin skeleton (with a nonreduced D ring; see Fig. 1) from the precursors DVP and MVP. In canonical chlorophyll biosynthesis pathways, the conversion of the porphyrin skeleton into a chlorin skeleton is catalyzed by protochlorophyllide oxidoreductases (PORs) (*23*) (Fig. 1). We hypothesized that the porphyrin skeleton retention in Chl c_2 and c_1 may depend on the substrate specificities of CHLC and PORs—i.e., the conversion of DVP and MVP into Chl c_2 and c_1 by CHLC may preclude further conversion by PORs, and conversely, the conversion of DVP into divinyl chlorophyllide *a* (DVC) may prevent the action of CHLC (Fig. 3C).

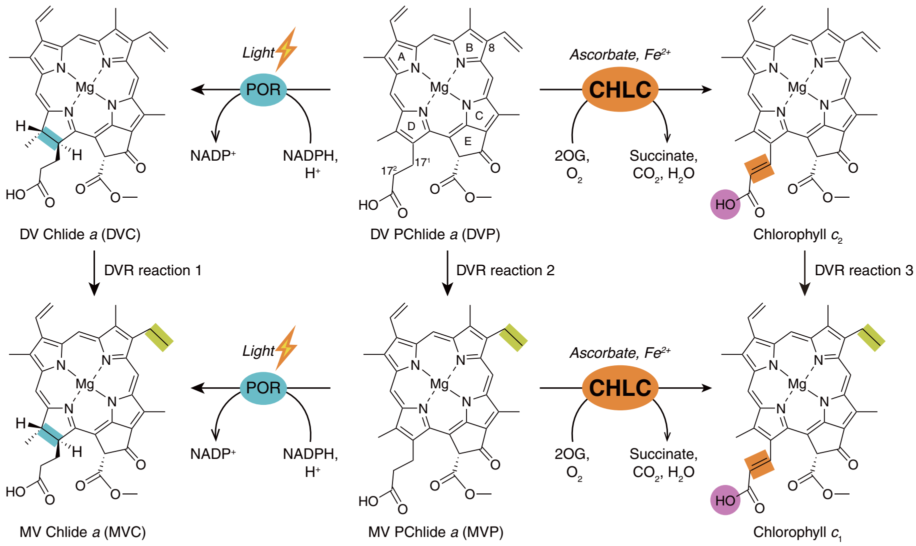


Fig. 1. An overview of the key steps of Chl *c* biosynthesis in *P. tricornutum*. Details of the steps from glutamate to DVP and the steps from monovinyl chlorophyllide *a* (MVC) to Chl *a* are skipped for simplicity. Carbon atoms 8, 17^1, and 17^2 and the five rings are labeled on DVP. Enzymes (shown as ovals) include CHLC (this study), divinyl reductase (DVR), and POR. Intermediates and cofactors include 2OG, DVC, DVP, MVC, and MVP. Shaded blocks indicate structural features generated by the corresponding enzymes (brown, CHLC; blue, POR; green, DVR). The circles in magenta highlight the protonation site absent in other chlorophylls. $NADP^+$, nicotinamide adenine dinucleotide phosphate; NADPH, reduced form of $NADP^+$.

We prepared DVC as a substrate for CHLC using in vitro reactions catalyzed by a characterized plant POR protein (*Arabidopsis thaliana* PORB) (*24*). We successfully obtained DVC based on the product's absorbance spectrum, including a prominent, red-shifted Qy band at 667.3 nm (fig. S4; see materials and methods for purification details). If CHLC could act on DVC, we expected that the addition of CHLC would generate a product with a strong Qy band as well but with an earlier elution profile compared with that of DVC based on the differences in elution between DVP (MVP) and Chl c_2 (Chl c_1) on our HPLC systems (Fig. 2H and Fig. 3A). We did not find such a product. All the contaminant peaks associated with our DVC preparation or generated by CHLC had low absorbance in the red region, characteristic of compounds containing a porphyrin skeleton (Fig. 3, D and E). These results suggested that DVC could not act as a substrate for CHLC.

Previously, land plant PORs have been found not to act on Chl c_2 and Chl c_1 (*25*). Assuming that PORs in *P. tricornutum* and plants have similar substrate specificities, the data suggest that CHLC and PORs both act on DVP in a mutually exclusive manner, and this can explain how the porphyrin skeleton is retained in Chl *c* (Fig. 3C).

CHLC is chloroplast localized and is required for normal physiology

We took advantage of the *chlc* mutants to understand the function of the CHLC protein and Chl *c*. Proteomic comparison of *chlc-1* with WT cells revealed multiple processes affected in the mutant, reflected as altered abundance of many proteins, although with a modest false discovery rate (FDR). The changes include lower relative abundance of some of the chlorophyll-binding proteins in the Lhcf family, lower abundance of a flavodoxin, and higher abundance of some of the reaction center subunits of the two photosystems (data S4). Notably, a few proteins annotated as heat shock protein Hsp20 isoforms were also up-regulated in the mutant. Previously, hyperaccumulation of Hsp20 family members has been observed in dark-treated *P. tricornutum* cells (*26*). Thus, our results are in line with Chl *c* playing an important role in photosynthesis.

The *chlc-1* mutant showed a decreased content of Chl *a* on a per-cell basis compared with the wild type, but the decrease of fucoxanthin was more prominent (fig. S5A), which explains the greenish appearance of the mutant (Fig. 2D). Consistently, the whole-cell absorbance spectra showed *chlc* mutants to be deficient in relative absorbance around 450 to 580 nm, around the absorbance peaks of Chl *c* and fucoxanthin (*6*) (fig. S5B). This absorption defect led to an actual effect on photosynthetic energy harvesting based on chlorophyll fluorescence measurements at 77 K—the *chlc-1* mutant had lower amplitudes compared with the wild type, with excitation wavelengths from 450 to 550 nm (fig. S5C).

Consistent with a function of Chl *c* and fucoxanthin in light harvesting, expression of *CHLC* in the wild type decreased after being

transferred from low-light to high-light conditions (fig. S5D); as controls, the chlorophyll biosynthesis gene *CHLM* and an LHC gene *LHCR10* were down-regulated and up-regulated by high light, respectively, consistent with previous studies (*27*). *chlc* mutants showed growth defects under low light but not as prominently during the first several days under high light. The mutants did grow slower than the wild type in the later stage under high light, possibly as a result of the reduced light availability to cells caused by a higher cell density (fig. S5, E and F).

We used the DiatOmicBase (https://www.diatomicsbase.bio.ens.psl.eu/) page of *CHLC* to learn its behavior in previous genome-wide studies. CHLC was predicted as a chloroplast protein by ASAFind (*28*). We confirmed that the GFP signal from the CHLC-GFP fusion protein coincided with the chlorophyll autofluorescence in COMP1 cells (fig. S6). Besides high light, previous studies have shown that nitrogen limitation, a commonly occurring stress for phytoplankton, also lowers the expression of *CHLC* (*29*). Notably, H3K9/14Ac (active chromatin mark) and H3K4me2 histone modifications have been shown to occur at the *CHLC* locus (*30*). Under nitrogen limitation, both marks exhibited lower signals and upstream-shifted peaks (fig. S7). These results suggest epigenetic mechanisms to be involved in the regulation *CHLC* expression.

Phylogenetic survey of *CHLC* within ochrophytes suggests the presence of at least one distinct, yet-unknown Chl *c* synthase

We investigated the distribution of *CHLC* among representative ochrophyte algae (photosynthetic stramenopiles; fig. S1, Fig. 4A, and data S5). We noted that diatoms (Bacillariophyta), Bolidophyceae, Pelagophyceae, and Dictyophyceae all have both Chl *c* (*6*, *18*) and *CHLC* orthologs [E value cutoff 1E^{-20} (1×10^{-20})]. These algae are closely related and belong to the so-called Diatomista group, one of the two groups within ochrophytes (*31*).

By contrast, brown algae (Phaeophyceae), Raphidophyceae, and Synchromophyceae (not shown in Fig. 4A) that belong to the Chrysista

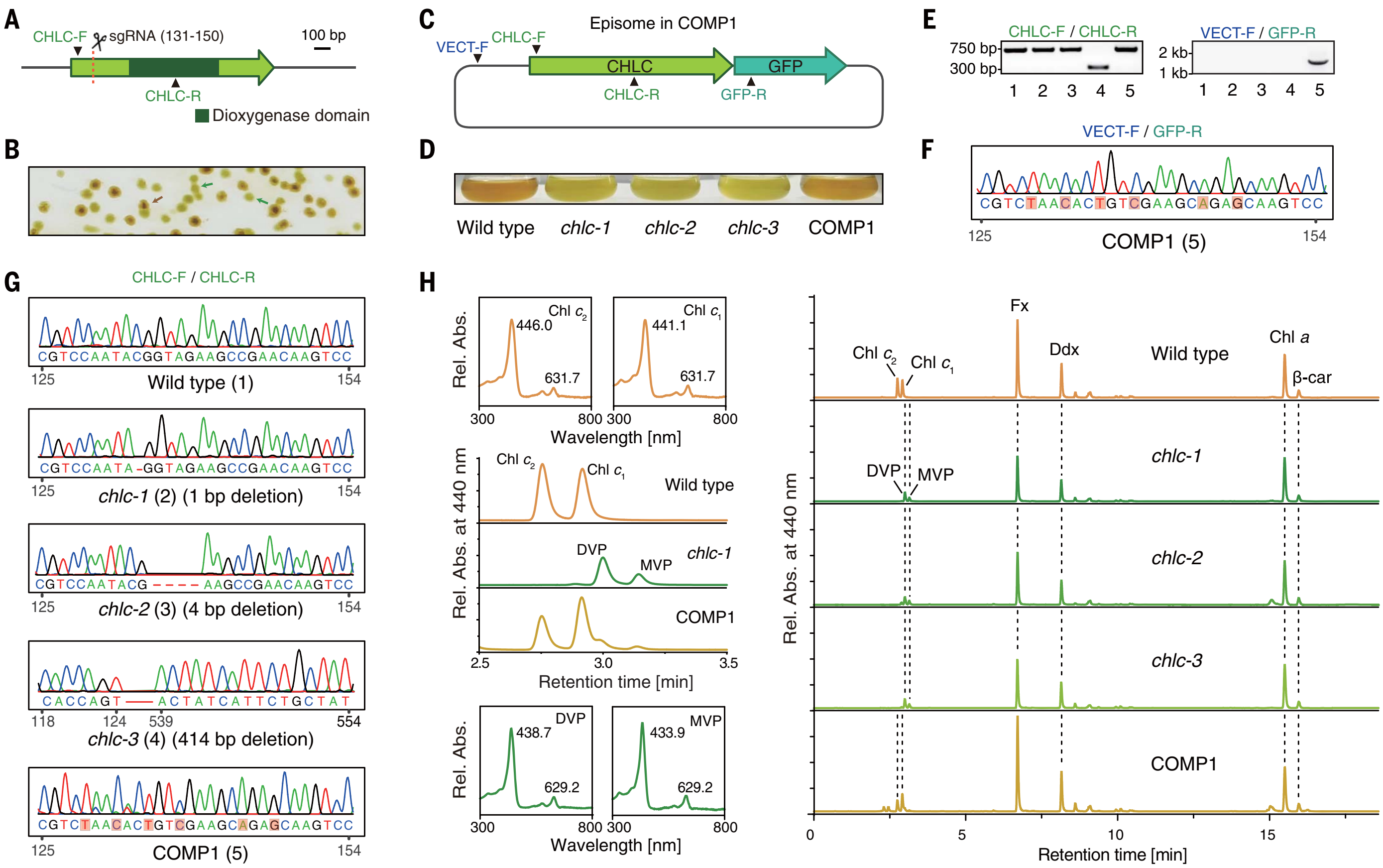


Fig. 2. *P. tricornutum* CHLC is required for the accumulation of Chl c_2 and c_1. (**A**) The intronless *CHLC* gene encodes a protein with a putative oxygenase domain. The schematic shows the positions of genotyping primers CHLC-F and CHLC-R and also the position of the single-guide RNA (sgRNA) used in CRISPR-Cas9 mutagenesis. (**B**) Part of an agar plate containing colonies from a transformation designed to disrupt *CHLC*. Green and brown arrows point to example transformants that we rated as green and brown clones. (**C**) A schematic representation of the additional, episomal copy of the *CHLC* gene present in the complemented COMP1 line. (**D**) A photograph of liquid cultures of the wild type, three mutants disrupted in *CHLC*, and the COMP1 line. (**E**) Agarose gel electrophoresis of polymerase chain reaction (PCR) products yielded from wild type, *chlc* mutants, and the COMP1 line using the two primer pairs shown in (A) and (C). Strains are numbered consistently across panels (E) to (G). bp, base pair. (**F**) Part of the sequencing results from the PCR product yielded by COMP1 using the primer pair CHLC-F–GFP-R; the size and sequence are consistent with an intact *CHLC-GFP* fusion gene. (**G**) Genotyping results of the *chlc* mutants and COMP1. Sequencing results from the PCR products in (E) are shown. Nucleotide positions in the coding sequence of *CHLC* are labeled. For COMP1, the sequence signals were mixed because both the mutated copy and the episomal copy on the episome were amplifiable using the CHLC-F–CHLC-R primer pair. In (F) and (G), synonymous substitutions in the episomal copy of *CHLC* are shown with a red shade. These were designed to avoid further editing of this episomal copy (see materials and methods). (**H**) HPLC analyses of pigment extracts of wild type, *chlc* mutants, and the COMP1 line. The traces were normalized against the height of the Chl *a* peak in each sample. The parts containing Chl c_2 and Chl c_1 peaks in wild type and COMP1 as well as that containing DVP and MVP peaks in *chlc-1* are magnified and shown on the left for clarity. The absorption spectra of these four peaks are presented on the left. Car, carotene; Ddx, diadinoxanthin; Fx, fucoxanthin; Rel. Abs., relative absorbance. For (D) and (H), representative results from no less than three replicates are shown.

group contain Chl *c* but not *CHLC* orthologs (*6*, *18*, *32*). In a recent study, we have reported that some of these Chrysista algae accumulate fucoxanthin as the major xanthophyll used for light harvesting, like diatoms do, but use a biosynthetic path different from that used by diatoms (*17*). Our findings in this work suggest a coevolution between the chlorophyll and carotenoid components of the light-harvesting antennae. It is possible that Diatomista and Chrysista algae each evolved Chl *c* and fucoxanthin biosynthesis separately. Alternatively, Chl *c* and fucoxanthin may have been present in the common ancestor, and one of the two lineages lost their biosynthetic pathways and then evolved different pathways. These hypotheses remain to be tested.

Among other Chrysista groups, Eustigmatophyceae do not synthesize Chl *c* and do not harbor any gene highly similar to *PtCHLC*, whereas Chrysophytes include species with or without Chl *c* (*6*, *18*, *33*); no *CHLC* ortholog was found in their genomes or transcriptomes. For Pinguiophyceae, a consensus has not been reached on whether they should be grouped into Diatomista or Chrysista (*31*, *34*). They contain both Chl *c* (*18*) and *CHLC* (data S5).

We caution that some of the sequence datasets used in our surveys are transcriptomes, and certain genes may have not been detected because of their low expression levels under the given conditions (data S5). Nevertheless, *CHLC* is consistently absent in the genomes and transcriptomes of Chrysista species examined, which increases our confidence in the presence of at least one additional synthase of Chl *c*.

CHLC is widely distributed within cryptophytes, haptophytes, and dinoflagellates and is the predominant Chl *c* synthase among known species

We then examined the occurrence of *CHLC* beyond ochrophytes. When PtCHLC was searched by BLAST against Tara Oceans Single-Cell and Metagenome Assembled Genomes sequences for marine eukaryotes (EUK-SMAGs) and the unassembled Marine Atlas of Tara Oceans Unigenes (MATOU) metatranscriptomics database (*35*–*37*), prevalent signals were observed across the sampling stations (fig. S8, A and C). The EUK-SMAGs hits were from various groups, especially haptophytes and ochrophytes (fig. S8B).

To obtain evolutionary insights into the CHLC family, we combined two approaches to search for its relatives. We (i) surveyed major groups of photosynthetic organisms with existing genomes or transcriptomes and (ii) performed an unbiased BLAST search in the UniProt database that includes heterotrophic organisms [E value cutoff 1E^{-5} (1×10^{-5})]. *CHLC* orthologs with an E value smaller than 1E^{-20} ($<1 \times 10^{-20}$) were found in cryptophytes, haptophytes, and many dinoflagellates in addition to Diatomista

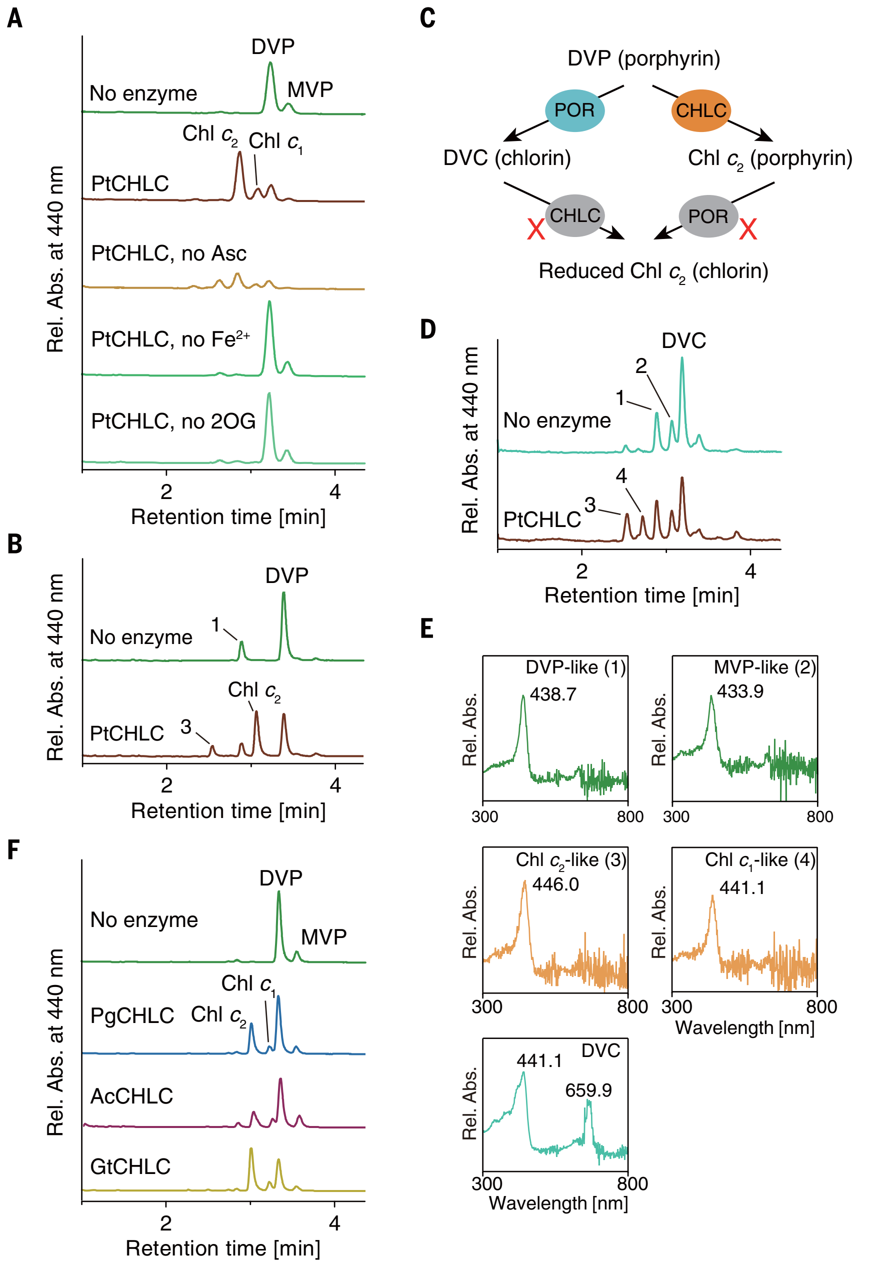


Fig. 3. Analysis of CHLC activities in vitro. Ascorbate (Asc), Fe^{2+}, and 2OG were added into the reaction mixtures as described in the materials and methods, unless specified. (**A**) CHLC produced Chl c_2 and Chl c_1 from DVP and MVP. The activity required Fe^{2+} and 2OG and was enhanced by Asc. (**B**) CHLC produced Chl c_2 from DVP purified to be free of MVP. See below for hypotheses on the contaminating peak 1 and on peak 3 generated by CHLC. (**C**) A schematic representation of the proposed substrate specificities. Red crosses indicate conversions that would not occur as a result of substrate discrimination. (**D**) CHLC does not act on the product of POR (DVC) containing a chlorin skeleton instead of a porphyrin skeleton. The peaks that we labeled DVP-like (peak 1) or MVP-like (peak 2) are similar to DVP or MVP in their absorption peak wavelengths in the blue light region (see Fig. 2H) and could be their isomers yielded from the pigment preparation process (*50*). The peaks labeled Chl c_2–like (peak 3) and Chl c_1–like (peak 4) are also similar to the corresponding compounds in absorption properties (Fig. 2H). (**E**) The absorbance profiles of pigment peaks from the experiment in (D). (**F**) Enzymatic analyses of the recombinant proteins for the orthologs of PtCHLC. PgCHLC: *Phaeocystis globosa* (haptophyte) Phaglo1-9344; AcCHLC: *Amphidinium carterae* (dinoflagellate) HBNO01055463.1; GtCHLC: *Guillardia theta* (cryptophyte) Guith1-93254. See data S5 for information on the genome or transcriptome database.

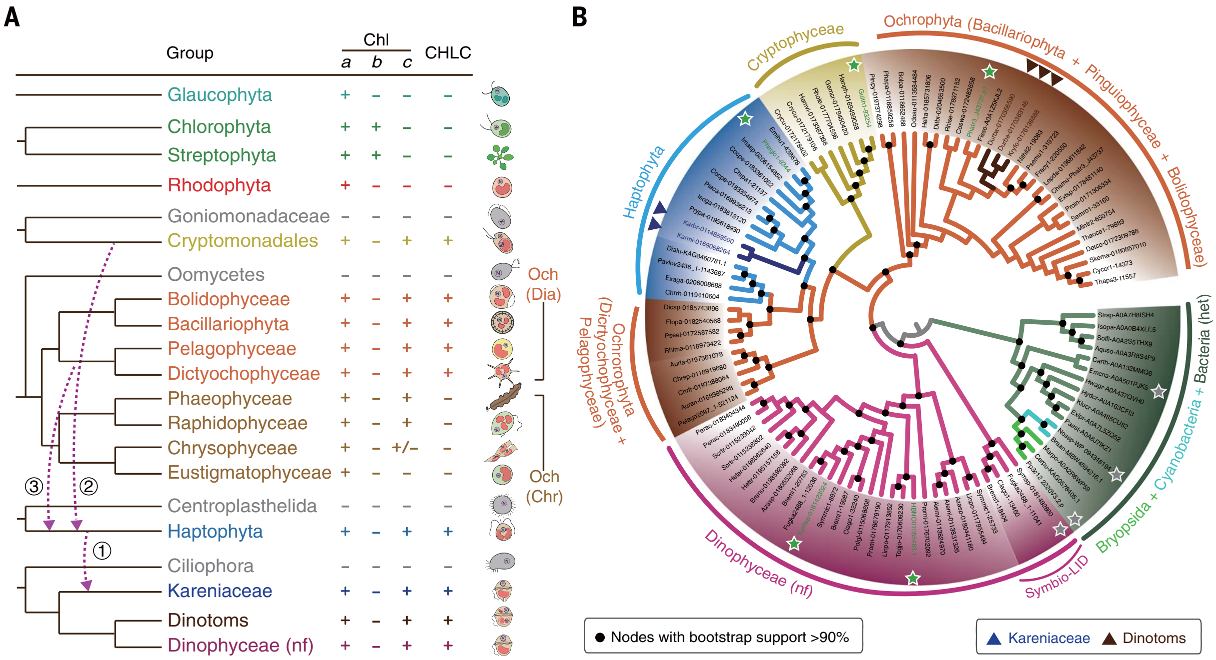


Fig. 4. Occurrence of CHLC and related proteins across the red and green lineages of photosynthetic eukaryotes. (**A**) The chlorophyll species and the presence or absence of CHLC orthologs in selected groups of photosynthetic eukaryotes and their plastid-less relatives (shown in gray fonts). Dinophyceae (nf), dinoflagellates that do not contain fucoxanthin; Och (Chr), Chrysista ochrophytes; Och (Dia), Diatomista ochrophytes. Chl *c* in this panel includes Chl c_2, Chl c_1, Chl c_3, Chl $c_{CS\text{-}170}$, and their esters with lipophilic molecules (*18*). Pigment compositions are displayed according to recent studies or reviews (*6*, *18*). The three arrows indicate three endosymbiotic events that were mentioned in the text. (**B**) A maximum-likelihood phylogenetic tree of the CHLC family. The green-filled stars indicate PtCHLCs and functionally validated CHLC orthologs, whereas the gray-filled stars indicate proteins unable to generate Chl *c* based on in vitro assays (Fig. 3F and fig. S10). Please see the text for a discussion of outgroup selection. Symbio-LID refers to Symbiodiniaceae low-identity blast hits. Protein names are edited to include letters from the genus or species names. For the original IDs, see data S5.

ochrophytes but were not found in glaucophytes, green algae, land plants, and chromerids. Two cyanobacterial genes, several genes from heterotrophic bacteria, and a number of genes from photosynthetic eukaryotes were deteced with a larger E value or lower identity compared with *PtCHLC* (Fig. 4A and data S5). *CHLC* orthologs were also absent in red algae as well as the plastid-less relatives of Chl *c*–containing algae, such as Goniomonadaceae, Centroplasthelida, Ciliophora, Oomycota, and Actinophryidae. This occurrence pattern corresponds well with the ability to synthesize Chl *c* among these groups (*6*, *18*) (fig. S1A and Fig. 4A).

We experimentally tested the enzymatic activity of three representative CHLC orthologs—from the cryptophyte *Guillardia theta* (Gt), the haptophyte *Phaeocystis globosa* (Pg), and the dinoflagellate *Amphidinium carterae* (Ac). All three proteins converted DVP and MVP into Chl c_2 and Chl c_1 in vitro (Fig. 3F), demonstrating a conserved function of the CHLC family among the four CASH lineages.

To visualize the phylogenetic relationship of CHLC orthologs within the four CASH lineages, we constructed a phylogenetic tree (Fig. 4B) with these proteins using blast hits from cyanobacteria, mosses, and heterotrophic bacteria as an outgroup; all these organisms lack Chl *c* and are well known to be distant from the CASH lineages. Among the high-identity (>40%) hits of CHLC (Fig. 4B and data S5), there appear two main clades: One was composed of dinoflagellates with a fucoxanthin-less plastid, and the other was composed of ochrophytes (Diatomista), haptophytes, and cryptophytes. The latter also contained dinoflagellates with a fucoxanthin plastid. Kareniaceae species are a group of dinoflagellates harboring fucoxanthin plastids that originated from engulfed haptophytes to replace the ancestor peridinin plastid (*38*) (arrow 1 in Fig. 4A). Their *CHLC* genes are also related to the haptophyte ones (Fig. 4B), which suggests a late plastid origin. Haptophytes themselves were hypothesized to contain a plastid (and its genome) from an engulfed cryptophyte (*39*) (arrow 2 in Fig. 4A). In addition, multiple haptophyte proteins predicted with a chloroplast localization seem to be of an ochrophyte origin, and one existing explanation is that an ochrophyte related to pelagophytes and dictyochophytes was once engulfed as an earlier plastid of haptophytes (*40*–*42*) (arrow 3 in Fig. 4A). If the latter is true, *CHLC* orthologs in haptophyte nuclear genomes may be of this origin, considering the close clustering between haptophytes and this group of ochrophytes in the tree. Many MATOU sequences with unknown species information cluster with known ochrophyte and haptophyte species (fig. S8D), which suggests that the relationship between ochrophyte and haptophyte CHLC orthologs can be better resolved when more species can be cultured for sequencing or are sequenced at the single-cell level.

A group of dinoflagellates (*Symbiodinium* sp. CCMP421, *Breviolum minutum*, *Cladocopium goreaui*, and *Fugacium kawagutii*) of the family Symbiodiniaceae (order Suessiales) had two or three blast hits with an E value smaller than 1E^{-20} (<1×10^{-20}), but for each species, one of the hits had a low identity (below 30%; data S5). These Symbiodiniaceae low-identity proteins (referred to as Symbio-LID hereafter) form a branch separate from the high-identity CHLC orthologs (Fig. 4B). For *Symbiodinium* sp. CCMP421, both the high-identity hit CAMPEP_0181423024 (labeled Symsp_0181423024 in Fig. 4B) and the low-identity CAMPEP_0181492860 were functionally tested in vitro. Only the former (referred to as SsCHLC hereafter) generated Chl *c* (fig. S10). This may be related to the lack of large stretches of amino acids (corresponding to PtCHLC positions 255 to 279 and 321 to 336) and changes of residues around the "HXD/E" motif in the active site (*43*) in Symbio-LID compared with active CHLC members (fig. S11).

To probe into the emergence of the activity of CHLC in evolution, we constructed an unrooted phylogenetic tree using blast hits satisfying the E value cutoff of 1E^{-5} (1×10^{-5}) from groups beyond CASH lineages (data S5, fig. S9, and materials and methods). Notably, the CHLC orthologs with a high identity (larger than 40%) in the four CASH lineages still form a monophyletic group. Low-identity hits from bacteria, Hexanauplia, Chlorarachniophyceae, Bryopsida, and algae of the CASH lineages are intermixed, obscuring the origin of the CHLC family. For those proteins not closely clustered with PtCHLC, we tested the activity of hits from a cyanobacterium, a proteobacterium, and a moss as well as an algal hit (*P. tricornutum* Phatr3_EG01494). They failed to produce Chl *c* in vitro (fig. S10), although it is possible that they function as dioxygenases on different compounds. Our sequence alignment revealed multiple residues, especially those around the HXD/E motif, to be conserved among and specific for the active CHLC members but different in bacterial hits (fig. S12). Some of these residues may have played a role in the emergence of the Chl *c* synthase activity—e.g., by facilitating the alteration of substrate specificity.

Discussion

Our discovery of the broadly conserved CHLC enzyme resolves a major outstanding question in photosynthesis. We demonstrate that CHLC is a 2OG-dependent dioxygenase that synthesizes Chl *c* from common chlorophyll biosynthesis intermediates. Both Chl *a* and Chl *c* are derivatives of the DVP precursor, and Chl *a* requires more steps to be synthesized (Fig. 1). Yet, the adoption of a dioxygenase for the production of Chl *c* suggests its evolutionary emergence to be after that of Chl *a* because oxygenic phototrophs almost universally use Chl *a* in the photosystem reaction centers to enable water oxidation (*44*). However, we can not exclude the possibility that the unidentified Chrysista Chl *c* synthase(s) are oxygen independent, considering that many known steps in tetrapyrrole biosynthesis can be catalyzed by evolutionarily unrelated oxygen-dependent or oxygen-independent enzymes (*12*). To identify the Chrysista Chl *c* synthase (s), one can again select candidate genes predicted to encode oxidoreductases by using transcriptomic datasets or comparative genomic approaches. Whereas genetic approaches are less advanced for Chrysista species compared with *P. tricornutum*, the *chlc* mutants from our study would allow a facile, heterologous expression-based candidate screen based on reversion of their green color.

CHLC is conserved among Diatomista ochrophytes, cryptophyte, haptophytes, and dinoflagellates but is absent in red algal plastid and nuclear genomes (data S5). It is only distantly related to any gene in organisms without Chl *c*, whether photosynthetic or not. How the Chl *c* synthase activity first came into existence thus remains a mystery. Regarding the inheritance of *CHLC* across CASH lineages, the potential early and late plastid origins of haptophyte and Kareniaceae *CHLC* genes provide support for the hypothesis that plastids serve as "shopping bags" for the evolution of the host genome (*45*). The observation that cryptophytes branch within ochrophytes in the tree of the CHLC family (Fig. 4B) is intriguing, considering that cryptophytes were hypothesized to be the earliest lineage among CASH algae in serial endosymbiosis models (*39*, *41*, *42*) (fig. S1C). This branching topology may be a phylogenetic artifact but could also have resulted from a gene transfer from an ochrophyte to an ancestor of extant cryptophytes.

Our *chlc* mutants allow for the investigation of the consequences of the loss of Chl *c*. Chl *c* likely emerged as a replacement to phycobilisomes for light harvesting with a lower nitrogen, sulfur, and photon cost (*46*). Because red algae already contained chlorophyll-binding LHC proteins (*9*), only modest changes were required to accommodate this pigment. In addition, Chl *c* may have other roles related to its structural uniqueness. Most Chl *c* species are not esterified to a hydrophobic phytol tail, leaving a free carboxyl end that can be reversibly protonated, which has recently been hypothesized to permit its versatile functions in both harvesting of light and protection from excess irradiation in a condition-dependent manner (*47*). Furthermore, the lack of the phytol group allows more fucoxanthin to bind to the LHC proteins (*48*). Moreover, Chl *c* was proposed to facilitate energy transfer from blue-green light–harvesting carotenoids to Chl *a* in the light-harvesting apparatuses (*1*), and it has now been shown by structural approaches that Chl *c* can interact with both fucoxanthin and Chl *a* in *P. tricornutum* (*48*). Our results of 77 K chlorophyll fluorescence spectra and growth curves under limiting light support the role of Chl *c* in light harvesting, and the defects can be partially attributed to the secondary effect on fucoxanthin or even Chl *a* accumulation (fig. S5). Regarding the lowered relative fucoxanthin content in *chlc* mutants, whether it resulted from biosynthetic coordination of the two pigments or whether lack of Chl *c* destabilizes the fucoxanthin-chlorophyll protein complexes (*48*) remains to be investigated.

It should be noted that algal groups in the four CASH lineages are different from each other in the carotenoid component in the Chl *c*–containing light-harvesting antennae; for instance, cryptophytes and many dinoflagellates use alloxanthin and peridinin instead of fucoxanthin (*6*). It would not be surprising if there were distinct aspects in the structural and functional roles of Chl *c* between different lineages. The identificaion of *CHLC* will allow researchers to directly target its orthologs in cryptophyte, haptophyte, or dinoflagellate species to obtain mutants devoid of Chl *c*, without a need to screen candidates, once genetic tools for these algae become available. Results from these efforts will shed light on whether Chl *c* exerts conserved functions across lineages beyond its light-harvesting role.

Our discovery of a Chl *c* synthase will facilitate synthetic biology efforts to synthesize this pigment in other phototrophs for improvements of their light-harvesting capability. Algae of the *Nannochloropsis* genus (Eustigmatophyceae) offer an attractive chassis for such attempts because they are industrially useful and genetically tractable but lack accessory chlorophylls (*49*). Additionally, their ochrophyte identity may facilitate correct targeting of the CHLC enzyme.

REFERENCES AND NOTES

1. A. W. D. Larkum, in *Chlorophylls and Bacteriochlorophylls: Biochemistry, Biophysics, Functions and Applications*, B. Grimm, R. J. Porra, W. Rüdiger, H. Scheer, Eds. (Springer, 2006), pp. 261–282.
2. G. G. StokesII, *Proc. R. Soc.* **13**, 144–145 (1864).
3. M. Chen *et al*., *Science* **329**, 1318–1319 (2010).
4. M. Y. Ho, G. Shen, D. P. Canniffe, C. Zhao, D. A. Bryant, *Science* **353**, aaf9178 (2016).
5. D. Baurain *et al*., *Mol. Biol. Evol.* **27**, 1698–1709 (2010).
6. C. Büchel, *Biochim. Biophys. Acta Bioenerg.* **1861**, 148027 (2020).
7. C. B. Field, M. J. Behrenfeld, J. T. Randerson, P. Falkowski, *Science* **281**, 237–240 (1998).
8. J. M. Archibald, *Curr. Biol.* **19**, R81–R88 (2009).
9. B. R. Green, *Photosynth. Res.* **107**, 103–115 (2011).
10. S. W. Jeffrey, *J. Phycol.* **12**, 349–354 (1976).
11. M. Nymark, A. K. Sharma, T. Sparstad, A. M. Bones, P. Winge, *Sci. Rep.* **6**, 24951 (2016).
12. D. A. Bryant, C. N. Hunter, M. J. Warren, *J. Biol. Chem.* **295**, 6888–6925 (2020).
13. S. I. Beale, in *Chloroplast Biogenesis*, N. R. Baker, J. Barber, Eds. (Elsevier, 1984), pp. 133–205.
14. O. Ait-Mohamed *et al*., *Front. Plant Sci.* **11**, 590949 (2020).
15. S. J. Mihalik, A. M. Rainville, P. A. Watkins, *Eur. J. Biochem.* **232**, 545–551 (1995).
16. M. S. Islam, T. M. Leissing, R. Chowdhury, R. J. Hopkinson, C. J. Schofield, *Annu. Rev. Biochem.* **87**, 585–620 (2018).
17. Y. Bai *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2203708119 (2022).
18. M. Zapata, J. L. Garrido, S. W. Jeffrey, in *Chlorophylls and Bacteriochlorophylls: Biochemistry, Biophysics, Functions and Applications*, B. Grimm, R. J. Porra, W. Rüdiger, H. Scheer, Eds. (Springer, 2006), pp. 39–53.
19. A. Gavalás-Olea, J. L. Garrido, P. Riobó, S. Álvarez, B. Vaz, *Curr. Org. Chem.* **22**, 836–841 (2018).
20. N. P. Dunham *et al*., *J. Am. Chem. Soc.* **140**, 7116–7126 (2018).
21. W. C. Chang *et al*., *Science* **343**, 1140–1144 (2014).
22. R. C. Dougherty, H. H. Strain, W. A. Svec, R. A. Uphaus, J. J. Katz, *J. Am. Chem. Soc.* **92**, 2826–2833 (1970).
23. W. Rüdiger, in *Chlorophylls and Bacteriochlorophylls: Biochemistry, Biophysics, Functions and Applications*, B. Grimm, R. J. Porra, W. Rüdiger, H. Scheer, Eds. (Springer, 2006), pp. 189–200.
24. M. Gabruk *et al*., *PLOS ONE* **10**, e0116990 (2015).
25. H. Klement, M. Helfrich, U. Oster, S. Schoch, W. Rüdiger, *Eur. J. Biochem.* **265**, 862–874 (1999).
26. X. Bai *et al*., *Sci. Rep.* **6**, 25494 (2016).
27. M. Nymark *et al*., *PLOS ONE* **4**, e7743 (2009).
28. A. Gruber, G. Rocap, P. G. Kroth, E. V. Armbrust, T. Mock, *Plant J.* **81**, 519–528 (2015).
29. O. Levitan *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 412–417 (2015).
30. A. Veluchamy *et al*., *Genome Biol.* **16**, 102 (2015).
31. S. M. Adl *et al*., *J. Eukaryot. Microbiol.* **66**, 4–119 (2019).

32. M. Schmidt *et al*., *Protist* **163**, 544–559 (2012).
33. C. Grevby, C. Sundqvist, *J. Plant Physiol.* **140**, 414–420 (1992).
34. A. Di Franco, D. Baurain, G. Glöckner, M. Melkonian, H. Philippe, *Mol. Biol. Evol.* **39**, msab300 (2022).
35. Q. Carradec *et al*., *Nat. Commun.* **9**, 373 (2018).
36. T. O. Delmont *et al*., *Cell Genomics* **2**, 100123 (2022).
37. C. Vernette *et al*., *Nucleic Acids Res.* **50**, W516–W526 (2022).
38. T. Tengs *et al*., *Mol. Biol. Evol.* **17**, 718–729 (2000).
39. A. Bodył, J. W. Stiller, P. Mackiewicz, *Trends Ecol. Evol.* **24**, 119–121 (2009).
40. R. G. Dorrell *et al*., *eLife* **6**, e23717 (2017).
41. J. W. Stiller *et al*., *Nat. Commun.* **5**, 5764 (2014).
42. J. F. H. Strassert, I. Irisarri, T. A. Williams, F. Burki, *Nat. Commun.* **12**, 1879 (2021).
43. I. J. Clifton *et al*., *J. Inorg. Biochem.* **100**, 644–669 (2006).
44. L. O. Björn, G. C. Papageorgiou, R. E. Blankenship, Govindjee, *Photosynth. Res.* **99**, 85–98 (2009).
45. A. W. Larkum, P. J. Lockhart, C. J. Howe, *Trends Plant Sci.* **12**, 189–195 (2007).
46. J. A. Raven, *New Phytol.* **98**, 593–625 (1984).
47. M. Tsujimura, M. Sugano, H. Ishikita, K. Saito, *J. Phys. Chem. B* **127**, 505–513 (2023).
48. W. Wang *et al*., *Science* **363**, eaav0365 (2019).
49. E. Poliner, E. M. Farré, C. Benning, *Plant Cell Rep.* **37**, 1383–1399 (2018).
50. H. Scheer, in *Chlorophylls and Bacteriochlorophylls: Biochemistry, Biophysics, Functions and Applications*, B. Grimm, R. J. Porra, W. Rüdiger, H. Scheer, Eds. (Springer, 2006), pp. 1–26.
51. Y. Jiang, Phylogenetic tree files on PHYH, *Zenodo* (2023); https://zenodo.org/record/8153423.

ACKNOWLEDGMENTS

The authors thank Z. Mao (Westlake University) and J. Yang (Westlake University) for help in protein purification; T. You (Westlake University) for proofreading the manuscript; and Y. Zhen (Westlake University), L. Zhang (Westlake University), Y. Chen (Westlake University), Z. Chen (Westlake University), M. Fan (Westlake University), K. Guo (Westlake University), R. Liu (Westlake University), X. Zhao (Nanjing University), and W. Wang (Chinese Academy of Sciences) for helpful discussions. We also acknowledge the Instrumentation and Service Center for Molecular Sciences and the Biomedical Research Core Facilities at Westlake University for technical support. Our in silico investigation of the *CHLC* gene was facilitated by the DiatOmicBase website (https://www.diatomicsbase.bio.ens.psl.eu/). **Funding:** This study was supported by National Key R&D Program of China grant 2019YFA0906300 (X.L.), National Key R&D Program of China grant 2022YFC3401800 (T.C.), Zhejiang Natural Science Foundation Outstanding Youth grant LR20C020002 (X.L.), Zhejiang Province Key R&D Program grant 2023SDXHDX0002 (X.L.), National Natural Science Foundation of China grant 42206115 (T.C.), National Natural Science Foundation of China grant 42106114 (H.Z.), National Natural Science Foundation of China grant 32150022 (X.L.), and the Westlake Education Foundation (X.L.). **Author contributions:** Conceptualization: X.L. Investigation: Y.J., T.C., Y.Y., J.Z., and H.Z. Visualization: Y.J., T.C., and Y.Y. Funding acquisition: T.C., H.Z., and X.L. Writing – original draft: X.L. Writing – review & editing: Y.J., T.C., Y.Y., and H.Z. **Competing interests:** X.L., Y.J., and T.C. are inventors on the Chinese patent application titled "Application of the *CHLC* Gene of *Phaeodactylum tricornutum* and Its Encoded Protein in Chlorophyll *c* Synthesis" (Westlake University, filed 28 August 2023, application no. 202311096732.0). This patent covers the application of the *P. tricornutum CHLC* gene or its homologs in the enhancement of Chl *c* content or production of Chl *c* in plants, algae, and microbes. The other authors declare no competing interests. **Data and materials availability:** The *chlc-1*, *chlc-2*, and *chlc-3* mutant strains and the COMP1 strain can be requested from the corresponding author without a material transfer agreement requirement. The phylogenetic tree files and associated files containing the sequences and sequence alignments are available at Zenodo (*51*). All other data are available in the main text or the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg7921
Materials and Methods
Figs. S1 to S12
References (*52*–*79*)
MDAR Reproducibility Checklist
Data S1 to S5

Submitted 20 January 2023; accepted 30 August 2023
10.1126/science.adg7921

CATALYSIS

The oscillating Fischer-Tropsch reaction

Rui Zhang[1], Yong Wang[1,2]*, Pierre Gaspard,[3] Norbert Kruse[1,2,4]*

The mechanistic steps that underlie the formation of higher hydrocarbons in catalytic carbon monoxide (CO) hydrogenation at atmospheric pressure over cobalt-based catalysts (Fischer-Tropsch synthesis) have remained poorly understood. We reveal nonisothermal rate-and-selectivity oscillations that are self-sustained over extended periods of time (>24 hours) for a cobalt/cerium oxide catalyst with an atomic ratio of cobalt to cerium of 2:1 (Co_2Ce_1) at 220°C and equal partial pressures of the reactants. A microkinetic mechanism was used to generate rate-and-selectivity oscillations through forced temperature oscillations. Experimental and theoretical oscillations were in good agreement over an extended range of reactant pressure ratios. Additionally, phase portraits for hydrocarbon production were constructed that support the thermokinetic origin of our rate-and-selectivity oscillations.

The hydrogenation of carbon monoxide to chain-lengthened hydrocarbons was initially reported by BASF company in 1913, where the reaction occurred under hydrogen-deficient synthesis conditions (*1*). A decade later, Fischer and Tropsch (*2*) demonstrated hydrocarbon synthesis from stoichiometric mixtures with H_2/CO ratios of 2 (or even more H_2-rich mixtures) that were compatible with the "syngas" derived from coal gasification and adjusted by water gas shift. Mainly iron and cobalt metal particles on a suitable support material were found to be active catalysts. The underlying mechanistic steps of this polymerization-type surface reaction are still under debate more than 100 years later. No consensus has emerged on the chain-lengthening mechanism that leads to hydrocarbons and their functionalized derivatives, in part because of metal-support synergies that range from weak interactions to strong chemical bonding and compound formation. Also, structural and chemical reconstruction effects on both the surface and in the bulk of the catalytically active metals (*3*–*12*) further complicate mechanistic studies.

Recent experimental studies have demonstrated that an all-embracing reaction mechanism to produce paraffins, olefins, and oxygenates (mainly alcohols and aldehydes) is unlikely to exist. Besides metal-support interactions, process temperatures and variable syngas compositions at atmospheric and elevated pressure conditions affect the mechanistic pathways. In this context, we note that Fischer originally adopted the view of a "formate-type" intermediate (*2*) after the report by Christiansen (*13*) of the catalytic hydrogenation of methyl formate to methanol. However, Fischer soon discarded this possibility after realizing that excessive carbon deposition from CO over alkali-promoted iron catalysts would lead to the formation of metal carbides prone to undergoing hydrogenation to hydrocarbons. Much of the present-day research favors a "carbide-type" mechanism that involves C–C coupling of CH_x intermediates. However, experimental evidence has also been presented for a CO insertion mechanism under steady-state conditions of the Fischer-Tropsch (FT) process (*14*–*20*). This mechanism benefits from its ability to account for the formation of chain-lengthened oxygenates besides paraffins and olefins. Reducible oxides of metals, in particular, those containing manganese, have been demonstrated to

[1]Voiland School of Chemical Engineering and Bioengineering at Washington State University, Pullman, WA 99164, USA. [2]Institute for Integrated Catalysis, Pacific Northwest National Laboratory, Richland, WA 99332, USA. [3]Centre for Nonlinear Phenomena and Complex Systems, CP231, Université Libre de Bruxelles, B-1050 Brussels, Belgium. [4]Chemistry of Surfaces, Interfaces and Nanomaterials (ChemSIN), CP243, Université Libre de Bruxelles, B-1050 Brussels, Belgium.
*Corresponding author. Email: norbert.kruse@wsu.edu (N.K.); wang42@wsu.edu (Y.W.)

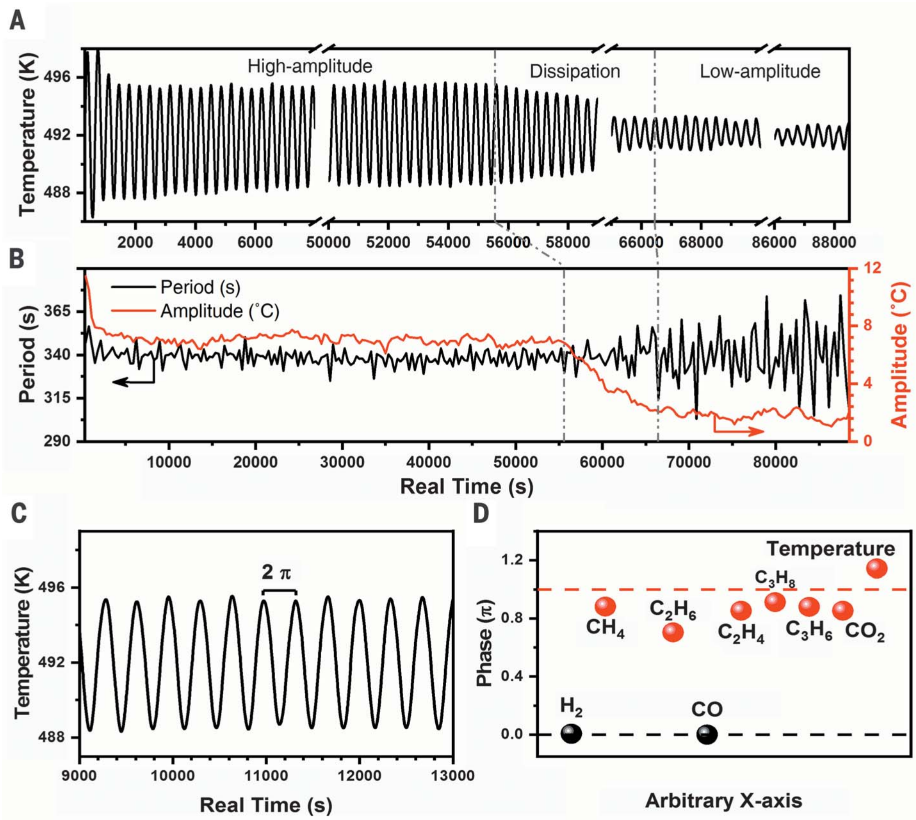


Fig. 1. Oscillatory behavior in the FT reaction. (**A** to **D**) Self-sustained long-term oscillations of the temperature over a Co/Ce–oxide catalyst of Co_2Ce_1 atomic composition at 220°C, 1 bar total pressure, and a H_2/CO ratio of 1/1. Oscillating temperatures (A) with periods of about 340 s and amplitudes of 7 Celsius degrees (for 16 hours) and 2 Celsius degrees (6 hours) (B) are observed along with a "dissipation" period in between and a short final dying-out. The harmonicity of the temperature oscillations is demonstrated in (C), and phases of temperature are related to those of the reactants and selected products after Fourier transformation in (D). Black and red dashed lines represent relative phases zero and π. C_4 products show the same behavior.

strongly promote the catalytic performance of cobalt-based catalysts (*9–12*, *21, 23–25*). Rate-and-selectivity hysteresis driven by a Co-to-Co_2C reaction-induced structural transformation has been reported for the high-pressure synthesis of chain-lengthened paraffins, olefins, alcohols, and aldehydes (*25*). Recent reports on using CeO_2 in Co-based catalyst formulations (*26*) highlighted the occurrence of a strong metal-support interaction (*27–29*) that led to the formation of patchy capping overlayers with enriched metal/CeO_x interface structures that serve as potential reaction sites (*30*). These insights challenged the authoritative view of Iglesia *et al.* (*31*) that hydrocarbon synthesis over Co-based catalysts is support independent.

Although time-dependent reaction kinetics to the point of hysteresis and oscillations were already demonstrated in the early 1970s for the CO oxidation over Pt-based catalysts (*32*, *33*), little is known so far about such kinetic instabilities in reaction networks as complicated as the catalytic CO hydrogenation. Because of possible heat- and mass-transfer limitations in solid catalysts, oscillations are usually nonisothermal. Indeed, Tsotsis *et al.* (*34*) observed temperature variations of up to 200 Celsius degrees in the CO hydrogenation over Fe/zeolite catalysts of poor thermal conductivity. In this work, we report periodic rate-and-selectivity oscillations in the FT reaction over a Co/Ce–oxide powder catalyst. Temperature amplitudes of several Celsius degrees were observed and were correlated to the periodic activity-and-selectivity changes of the Co/Ce–oxide catalyst. The experimental evidence allowed us to develop a mechanistic framework that shows how observed oscillatory reactions can be generated through periodic temperature forcing. We also present phase portraits for oscillatory reactant and product formation in agreement with the thermokinetic origin of our observations.

Experimental studies

We used a fixed-bed flow reactor designed to minimize concentration gradients of reactants and products to allow the reactor response to be monitored as a function of time when switching reactant flows between different compositions. In the present study, such a transient response was obtained after swiftly replacing a H_2 flow (for catalyst activation and to provide a metallic Co phase) in He with a reactive syngas mixture of H_2/CO in argon. Noble gases were used as balance and reference along with an independent flow of Ne that bypassed the reactor to allow molecular flows of reactants and products to be calculated, assuming negligible concentration gradients in the reactor and laminar flow in the pipes leading to the analytical devices. The experimental procedure that we adopted to generate rate-and-selectivity oscillations of the atmospheric-pressure FT reaction over Co/ceria is shown in fig. S1.

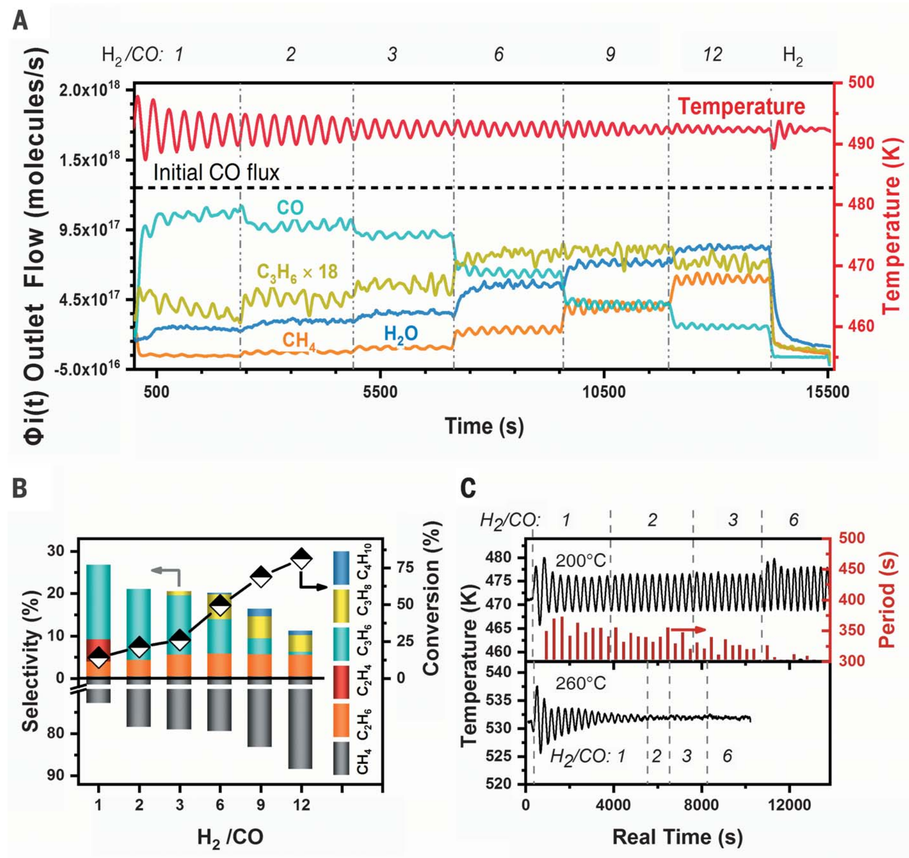


Fig. 2. Effect of H_2/CO ratio and temperature on the oscillatory behavior over a Co/Ce–oxide catalyst with Co_2Ce_1 atomic composition. (**A**) Evolution of temperature and rate-and-selectivity oscillations (as reactor outlet flows of reactants and products) as a function of the H_2/CO ratio at a nominal temperature of 220°C. The horizontal black dashed line represents the CO molecular flow into the reactor and serves to calculate time-averaged reaction rates. Dashed vertical lines distinguish reactant ratios between H_2/CO = 1 and 12. ϕi(t), reactor outlet flow. (**B**) CO conversion and product selectivities for varying H_2/CO ratios at 220°C. (**C**) Time evolution of oscillations at 200°C (top) and 260°C (bottom) for selected H_2/CO ratios. Red bars in the top graph reflect the variation of the oscillation period with time and the H_2/CO ratio.

After reducing conditions in H_2 switched to reactive conditions in H_2/CO, while the flow rate was kept constant, the catalytically active surface phase began to form. Such chemical conditioning always preceded the onset of oscillations and was associated with delay times of some hundreds of seconds, depending on the actual H_2/CO ratio (which varied from 1 to 12) and Co/Ce atomic ratios in Co/Ce–oxide catalysts (Co/Ce atomic ratios varied from 0.5 to 4). For the conditions shown in Fig. 1, long-term oscillations in temperature with periods of about 340 s and amplitudes of ~7 Celsius degrees were obtained for Co/Ce = 2 and H_2/CO = 1. After 15 hours, the amplitudes decreased to ~2 Celsius degrees (Fig. 1, A and B, and fig. S3). Such changes in amplitude were not abrupt but rather sluggish and occurred on timescales of ~1 hour. The origin of amplitude damping is not clear but could be associated with changes in the surface composition of the

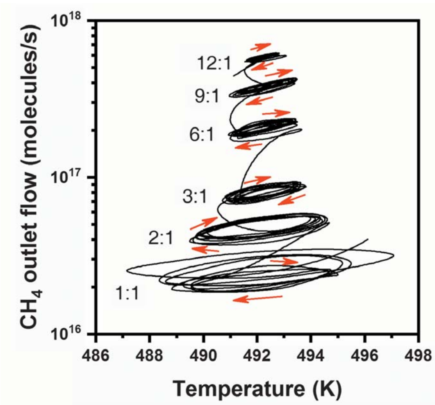


Fig. 3. Phase portrait of CH_4 selectivity versus temperature at different H_2/CO ratios. The experimental conditions are the same as in Fig. 2. Phase portraits, or phase-space trajectories, are constructed from extended time series of the oscillating product flows. Transitions to higher cycles are caused whenever the H_2/CO ratio is increased. The red arrows indicate the direction of time evolution.

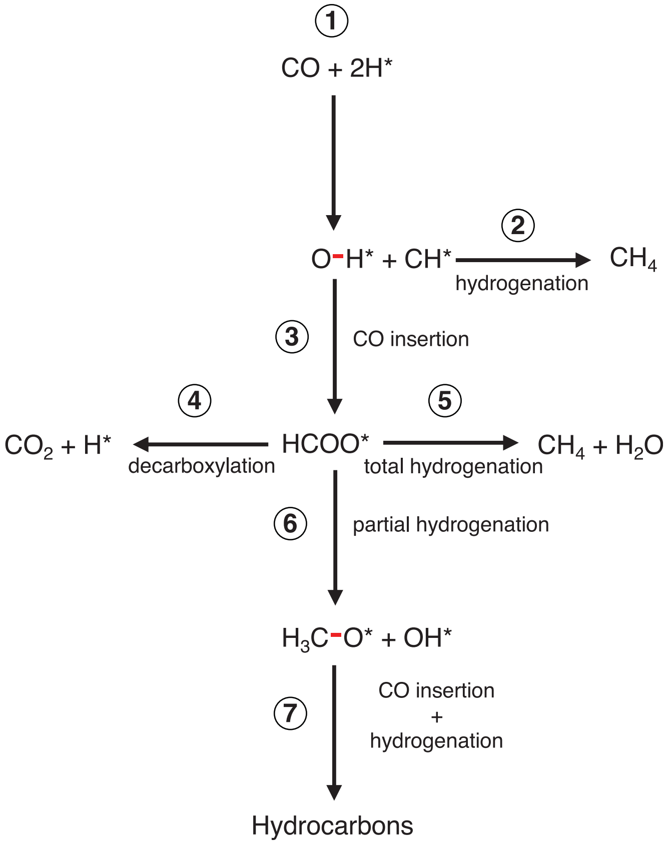


Fig. 4. Reduced reaction scheme that underlies theoretical modeling with forced temperature oscillations. Chain lengthening involves CO insertion into O–H or O–R bonds of surface hydroxy or alkoxy, respectively (critical bonds are highlighted in red). Methane formation occurs through either an "early" pathway via CH_x hydrogenation or a "late" pathway involving hydrogenation of surface formate. Only the late pathway develops oscillatory behavior, as do higher hydrocarbons. Generally, the reaction scheme lumps together several elementary steps to enable the mathematical treatment. Water gas shift is an integral part of the scheme. Possible surface reconstructions (due to Co-to-Co_2C or CeO_2-to-Ce_2O_3 chemical transformations) have not been considered, in accordance with x-ray diffraction results shown in fig. S7. The details of this reaction scheme along with the kinetic approach can be found in the supplementary materials.

catalyst from the adsorption of contaminating species.

Oscillating temperatures reflected periodic changes in the exothermicity of the reaction. In Fig. 1C, we show a selected time window to exemplify the harmonicity of the high-amplitude oscillations. Thermal coupling between remote catalyst particles had to be highly efficient to ensure the apparently concerted reaction behavior across the catalyst bed. In this context, no oscillatory behavior was observed after replacing Ar with He as the vector gas. The fast heat exchange between catalyst particles and gas phase that is enabled by the high thermal conductivity of He was lost after replacement with Ar (for details, see the supplementary materials).

Phase differences between temperature oscillations and reactants and products (Fig. 1D) showed that temperature maxima and minima correlated with enhanced and reduced product formation, respectively, whereas reactants were out of phase with both temperature and products. These interdependencies were also apparent from phase portraits, as will be shown for the case of methane formation.

Figure 2 demonstrates the oscillatory behavior as a function of the H_2/CO pressure ratio and temperature. Although this pressure ratio has been set to between 1 and 3 in most investigations of the FT reaction, we have chosen to extend it up to 12 after our previous studies over Co/Mn–oxide catalysts, which showed rate and selectivity hysteresis to endure a large excess of H_2 (*25*). Figure 2A demonstrates the exceedingly broad existence range of temperature and selected chem-species' oscillations as the H_2/CO ratio varies from 1 to 12. As this ratio is increased from low to high, the oscillation period decreased from ~340 to ~230 s and the amplitude decreased from 7 to 1 Celsius degree, respectively (fig. S3). Time-averaged reactor outlet flows decreased for CO and increased for all product species (exemplified for methane and propene in Fig. 2A), as expected for more-reducing reaction conditions.

Although the CO conversion increased from an initial value of 14% at $H_2/CO = 1$ to a final value of 81% at $H_2/CO = 12$, the product species exhibited a more-differentiated picture (Fig. 2B). Less-hydrogenated species, such as olefins, were formed preferentially at low H_2/CO ratios, as had been reported for promoted Co/Mn–oxide catalysts in the absence of long-term oscillations (*9*, *10*, *23*). According to high-resolution electron microscopy studies, a Co-to-Co_2C phase transformation occurred under these hydrogen-deficient conditions (*9*, *10*). Olefin formation, especially for ethylene, decreased at increasing H_2/CO ratios (Fig. 2B). The formation of methane, the most hydrogenated product, became dominant at higher H_2/CO ratios and reached ~90% selectivity at $H_2/CO = 12$. Although some chain lengthening to C_{2+} paraffins was retained, early chain termination was favored under these strongly reducing conditions.

Figure 2C illustrates that the oscillatory behavior described so far only occurred within a limited range of temperatures. At 200°C, regular oscillations were observed over a wide range of H_2/CO pressure ratios. Similar to observations made for a reaction temperature of 220°C (as mentioned earlier), oscillation periods decreased with increasing H_2/CO ratio. At 260°C, oscillations with fading amplitude were observed for $H_2/CO = 1$. The dissipation was then followed by the extinction of any oscillatory behavior at higher H_2/CO ratios. We note that Co/Ce–oxide catalysts with Co/Ce atomic compositions that were appreciably different from 2 (Co/Ce = 0.5, for example; see fig. S4) did not exhibit oscillatory behavior (fig. S4).

The phase-space trajectories of the methane outlet flow versus temperature are shown in Fig. 3. For each H_2/CO ratio, the trajectory formed a cycle with the orientation expected for a mechanism of thermokinetic instability (*35*, *36*). In this mechanism, a thermokinetic feedback caused by the high exothermicity of the FT reaction produced an increase of the temperature and an acceleration in the consumption of the adsorbed reactants, which then slowed the reaction and closed the cycle. In this way, the thermokinetic instability may trigger self-sustained oscillations (*37*), as observed in our research and described above.

Theoretical analysis

A deeper understanding of the intrinsically time-dependent reaction kinetics described thus far required a theoretical analysis of the reaction mechanism, the details of which are presented in the supplementary materials. A reduced scheme that summarizes the essential aspects of this mechanism is shown in Fig. 4. In particular, we assume hydrogen-assisted dissociation of CO (*15*, *38*) to provide O–H surface hydroxyl and partially hydrogenated carbon species, followed by hydrogenation of the latter to produce methane. Importantly, the O–H bond of OH undergoes an insertion reaction with CO to form surface formate-type species as a key surface reaction intermediate. A similar reaction scenario has also been envisaged for the methanol synthesis (*39*–*41*) over Cu-based catalysts. However, potential variations in the denticity of surface-bound formate, which could lead to different reactivities (*42*, *43*), have not been considered in our reaction scheme. The occurrence of CO insertion in the atmospheric FT reaction has been previously demonstrated in chemical transient studies using Co/MgO (*14*),

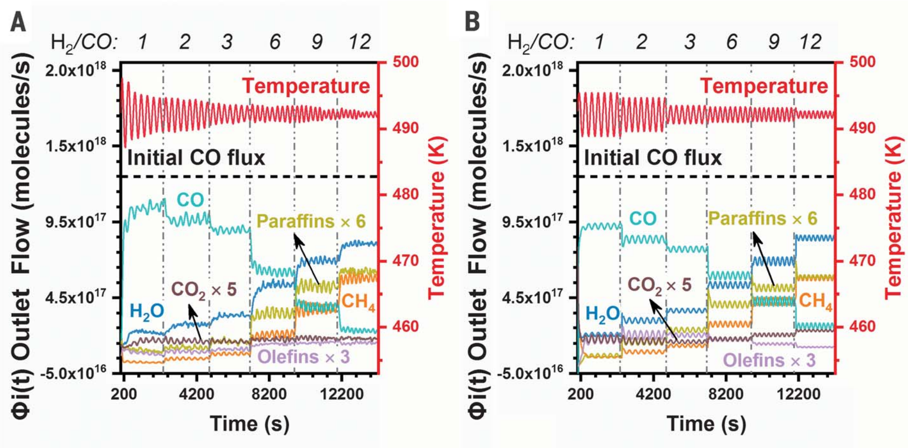


Fig. 5. Experimental and theoretical flow rates under conditions of self-sustained nonisothermal oscillations and forced temperature oscillations. (**A**) Time evolution of temperature and outlet flows under the same experimental conditions as in Fig. 2. (**B**) The same variables simulated with the theoretical kinetic model described in the supplementary materials for forced temperature oscillations. The outlet flow of CO is out of phase with temperature and products for the theoretical model in (B) as well as for the experimental observations in (A).

Co/Mn–oxide (*16*), and other catalyst compositions. The respective kinetic evidence for Co/Ce–oxide is also provided in the supplementary materials.

Chain lengthening involves surface hydrogenation of formate followed by CO insertion into the nascent surface O–R bond, where R is an alkyl group. Repetitive partial hydrogenation of surface carboxylate and CO insertion into the respective alkoxy O–R bond then provides chain lengthening. The reaction scheme also considers CO_2 formation from surface formate and may be completed to allow for olefin formation and hydrocarbon functionalization.

An essential feature of our reaction mechanism is the relatively late water rejection from the adsorbed layer. This result is in full agreement with results obtained from chemical transient kinetics for various catalyst systems (*14*, *16*, *44*). In the case of Co/Ce–oxide, fig. S2 shows delay times for water formation of several hundred seconds as compared with several seconds for methane formation. Reversible Ce^{4+}-to-Ce^{3+} reduction has not been considered in our reaction scheme nor has a Co-to-Co_2C transformation been accounted for. Including such structural changes in a complex kinetic reaction scheme like the FT reaction is challenging. Furthermore, no experimental evidence to support the occurrence of these transformations under atmospheric reaction conditions, as applied in our study, has been obtained (for details, see the supplementary materials).

Figure 5 compares the results of forced temperature oscillations with self-sustained nonisothermal oscillations as observed experimentally. The temperature dependence was assumed to arise mainly from Arrhenius factors associated with C–O bond breaking. The level of agreement between the two is deemed to be excellent. Time-averaged outlet flows of reactants and products deviate by about 20% (and usually much less) from each other over the entire range of H_2/CO ratios. We note that this level of agreement is strongly dependent on the values of the kinetic constants that are used (see supplementary materials). Earlier observations of large-amplitude (100° to 200°C) and very-long-period (usually 60 to 80 min) oscillations in the FT reaction at higher pressures (10 to 22 bar) did not provide deep mechanistic understanding (*34*, *45*). In addition to heat- and mass-transfer limitations, reversible deactivation through carbon deposition (*34*, *46*) and oxidation-reduction cycles were suggested to provide a chemical feedback for oscillations in temperature and either in-phase CO (*45*) or out-of-phase CO conversion (*34*).

Discussion

We report the discovery of rate-and-selectivity oscillations in the FT reaction over cobalt/ceria catalysts. The observations are modeled with a kinetic scheme based on the CO insertion mechanism, the thermal activation of C–O bond breaking, and periodic temperature forcing. Previous kinetic modeling studies that used the concept of forced oscillations to replicate self-sustained oscillatory behavior have generally been on much-less-complicated reaction networks. Ultimately, it would be of interest to generate self-sustained oscillatory behavior on the basis of a thermokinetic feedback mechanism without forcing the temperature to oscillate.

From an experimental point of view, the use of a nearly gradient-free microreactor turned out to be advantageous for our mean-field kinetic modeling. Although heat transport through the gas phase ensures "global" synchronization over the catalyst bed, local coverages at the surface of individual particles may vary greatly, thus causing surface diffusion and spatiotemporal self-organization. Such phenomena have been studied in great detail under isothermal and ultrahigh vacuum conditions with oriented single crystals (*47*) but escape detectability in ensemble-averaging analyses with support catalysts at ambient pressures. With regard to the FT reaction, kinetic studies under molecular flow conditions in ultrahigh vacuum are incongruous because of the strong pressure dependence related to the hydrogen reactant.

We anticipate that our approach will spawn further research to determine the appropriate parameters for triggering and analyzing kinetic instabilities of catalytic reactions that are as complicated as the FT reaction. This discovery can provide insight into the reaction pathways that lead to paraffins and olefins over a wide range of H_2/CO partial pressure ratios and benefit present FT research, up to the point of astrochemistry, where the reaction is being considered as a candidate for producing complex molecules from simpler ones (*48*–*50*). The results of our studies reinforce the concept of a dynamic catalyst design (*51*) to improve the performance in terms of rate and selectivity by using oscillatory reaction states. Concurrent theoretical analyses under thermokinetic conditions may guide the mechanistic understanding of reactions that are as complex as the FT synthesis.

REFERENCES AND NOTES

1. BASF, "Verfahren zur Darstellung von Kohlenwasserstoffen und deren Derivaten," German Patent 293,787 (1913).
2. F. Fischer, H. Tropsch, *Brennst. Chem* **4**, 276–285 (1923).
3. H. A. Bahr, V. Jessen, *Berichte der Dtsch. Chem. Gesellschaft A B Ser.* **63**, 2226–2237 (1930).
4. S. Weller, L. J. E. Hofer, R. B. Anderson, *J. Am. Chem. Soc.* **70**, 799–801 (1948).
5. R. B. Anderson, W. K. Hall, A. Krieg, B. Seligman, *J. Am. Chem. Soc.* **71**, 183–188 (1949).
6. J. A. Amelse, J. B. Butt, L. H. Schwartz, *J. Phys. Chem.* **82**, 558–563 (1978).
7. J. W. Niemantsverdriet, A. M. Van Der Kraan, W. L. Van Dijk, H. S. Van Der Baan, *J. Phys. Chem.* **84**, 3363–3370 (1980).
8. M. Claeys *et al.*, *J. Catal.* **318**, 193–202 (2014).
9. Y. Xiang, N. Kruse, *Nat. Commun.* **7**, 13058 (2016).
10. L. Zhong *et al.*, *Nature* **538**, 84–87 (2016).
11. Y. An *et al.*, *J. Catal.* **366**, 289–299 (2018).
12. Z. Zhao *et al.*, *ACS Catal.* **8**, 228–241 (2018).
13. J. A. Christiansen, "Method of producing methyl alcohol from alkyl formates," US Patent 1,302,011 (1919).
14. J. Schweicher, A. Bundhoo, N. Kruse, *J. Am. Chem. Soc.* **134**, 16135–16138 (2012).
15. M. Zhuo, K. F. Tan, A. Borgna, M. Saeys, *J. Phys. Chem. C* **113**, 8357–8365 (2009).
16. M. Athariboroujeny *et al.*, *ACS Catal.* **9**, 5603–5612 (2019).
17. R. A. van Santen, A. J. Markvoort, *ChemCatChem* **5**, 3384–3397 (2013).
18. B. Todic, W. Ma, G. Jacobs, B. H. Davis, D. B. Bukur, *Catal. Today* **228**, 32–39 (2014).
19. J. Hindermann, G. Hutchings, A. Kiennemann, *Catal. Rev. Sci. Eng.* **35**, 1–127 (1993).

20. S. Storsæter, D. Chen, A. Holmen, *Surf. Sci.* **600**, 2051–2063 (2006).
21. Y. Xiang *et al.*, *J. Am. Chem. Soc.* **135**, 7114–7117 (2013).
22. G. R. Johnson, S. Werner, A. T. Bell, *ACS Catal.* **5**, 5888–5903 (2015).
23. Z. Li *et al.*, *ACS Catal.* **7**, 3622–3631 (2017).
24. J. Paterson *et al.*, *ChemCatChem* **10**, 5154–5163 (2018).
25. Y. Xiang, L. Kovarik, N. Kruse, *Nat. Commun.* **10**, 3953 (2019).
26. M. K. Gnanamani *et al.*, *Appl. Catal. A Gen.* **393**, 17–23 (2011).
27. P. Wang *et al.*, *Ind. Eng. Chem. Res.* **61**, 3900–3909 (2022).
28. L. Spadaro *et al.*, *J. Catal.* **234**, 451–462 (2005).
29. Y. Liu *et al.*, *Fuel* **324**, 124518 (2022).
30. W. T. Figueiredo *et al.*, *ACS Appl. Nano Mater.* **2**, 2559–2573 (2019).
31. E. Iglesia, S. L. Soled, R. A. Fiato, *J. Catal.* **137**, 212–224 (1992).
32. P. Hugo, *Berichte der Bunsengesellschaft für Phys. Chemie* **74**, 121–127 (1970).
33. H. Beusch, P. Fieguth, E. Wicke, *Chemieingenieurtechnik* **44**, 445–451 (1972).
34. T. T. Tsotsis, V. U. S. Rao, L. M. Polinski, *AIChE J.* **28**, 847–851 (1982).
35. I. Y. Sal'nikov, *Dokl. Akad. Nauk SSSR* **60**, 405–408 (1948).
36. I. Y. Sal'nikov, *Zh. Fiz. Khim* **23**, 258–272 (1949).
37. I. V. Derevich, *Int. J. Heat Mass Transf.* **53**, 135–153 (2010).
38. M. Ojeda *et al.*, *J. Catal.* **272**, 287–297 (2010).
39. K. C. Waugh, *Catal. Today* **15**, 51–75 (1992).
40. F. C. Meunier, *Angew. Chem. Int. Ed.* **50**, 4053–4054 (2011).
41. Y. Yang, C. A. Mims, R. S. Disselkamp, C. H. F. Peden, C. T. Campbell, *Top. Catal.* **52**, 1440–1447 (2009).
42. D. Lorito, A. Paredes-Nunez, C. Mirodatos, Y. Schuurman, F. C. Meunier, *Catal. Today* **259**, 192–196 (2016).
43. J. Schweicher *et al.*, *J. Phys. Chem. C* **114**, 2248–2255 (2010).
44. A. Bundhoo, J. Schweicher, A. Frennet, N. Kruse, *J. Phys. Chem. C* **113**, 10731–10739 (2009).
45. L. Caldwell, in *Inst. Chem. Eng. Symp. Ser. 87* (1984), pp. 151–158.
46. M. M. Slin'ko, N. I. Jaeger, in *Oscillating Heterogeneous Catalytic Systems*, vol. 86 of *Studies in Surface Science and Catalysis* (Elsevier, 1994), chap. 4.
47. R. Imbihl, G. Ertl, *Chem. Rev.* **95**, 697–733 (1995).
48. M. H. Studier, R. Hayatsu, E. Anders, *Geochim. Cosmochim. Acta* **32**, 151–173 (1968).
49. E. Anders, R. Hayatsu, M. H. Studier, *Science* **182**, 781–790 (1973).
50. M. E. Kress, A. G. G. M. Tielens, *Meteorit. Planet. Sci.* **36**, 75–91 (2001).
51. M. Shetty *et al.*, *ACS Catal.* **10**, 12666–12695 (2020).

ACKNOWLEDGMENTS

Funding: R.Z. acknowledges financial support from Chambroad Chemical Industry Research Institute Co., Ltd. R.Z. and Y.W. acknowledge financial support from the US Department of Energy Basic Energy Sciences Catalysis Science program, grant DE-FG02-05ER15712. N.K. thanks the National Science Foundation for financial support under contract no. CBET-1438227. **Author contributions:** R.Z. performed the experiments and contributed to the drafting of the manuscript. Y.W. provided advice and contributed to the final drafting of the manuscript. P.G. performed the theoretical calculations and contributed to the drafting of the manuscript. N.K. initiated and supervised the project and also took the lead in drafting the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh8463
Materials and Methods
Supplementary Text
Figs. S1 to S7
Table S1
References (*52*–*55*)

Submitted 20 March 2023; accepted 18 August 2023
10.1126/science.adh8463

ANTHROPOLOGY

More than 10,000 pre-Columbian earthworks are still hidden throughout Amazonia

Vinicius Peripato *et al.*

Indigenous societies are known to have occupied the Amazon basin for more than 12,000 years, but the scale of their influence on Amazonian forests remains uncertain. We report the discovery, using LIDAR (light detection and ranging) information from across the basin, of 24 previously undetected pre-Columbian earthworks beneath the forest canopy. Modeled distribution and abundance of large-scale archaeological sites across Amazonia suggest that between 10,272 and 23,648 sites remain to be discovered and that most will be found in the southwest. We also identified 53 domesticated tree species significantly associated with earthwork occurrence probability, likely suggesting past management practices. Closed-canopy forests across Amazonia are likely to contain thousands of undiscovered archaeological sites around which pre-Columbian societies actively modified forests, a discovery that opens opportunities for better understanding the magnitude of ancient human influence on Amazonia and its current state.

During the pre-Columbian era, Amazonia was home to dense and complex societies throughout its vast forested area spanning 6.7 million km^2 (*1*). These ancient Indigenous societies had profound knowledge of earthmoving, riverine dynamics, soil enrichment, and plant and animal ecology, which allowed them to create domesticated landscapes that were more productive for humans (*2*–*4*). With earthmoving techniques, Indigenous peoples created a wide variety of earthworks (i.e., ring ditches, geoglyphs, ponds, and wells), mostly between 1500 and 500 years before present, with social, ceremonial, and defensive functions (*5*). Around these earthworks, they also managed hundreds of tree species, some of which show evidence of domestication (*6*–*9*), and effected long-lasting changes in forest composition (*10*–*13*). The scale and intensity of that landscape transformation remain unknown, in part because there has never been a comprehensive inventory of pre-Columbian sites across the basin.

Domesticated landscapes in Amazonia have mostly been discovered by means of evidence from on-the-ground surveys (*5*, *14*). Earthworks can be detected by orbital optical satellites with very high spatial resolution (*15*), but that technique is mostly suitable for deforested areas (*16*). Airborne light detection and ranging (LIDAR) data—a remote sensing technique that can map microtopography beneath the forest canopy—has substantially changed our understanding of the magnitude of pre-Columbian urbanism in Mesoamerica (*17*, *18*) and South America (*19*). Over the past decade, the use of LIDAR data has revealed the complexity of Mayan civilization by indicating a regionally integrated urban-rural community network in Mesoamerica (*17*). More recently, LIDAR enabled the detailed mapping of two monumental pre-Columbian settlements in an intensively domesticated landscape hidden under forest in southwestern Amazonia (*19*). Although Mesoamerican archaeological sites feature very different types of structures—stone construction as opposed to the use of earth, as in Amazonia—LIDAR technology has substantially improved our spatial understanding of archaeological sites in forested landscapes by enabling the visualization of ancient large-scale earthworks (*18*, *19*) beneath the forest canopy. Because deforestation in Amazonia has removed about 17% of the natural vegetation cover to date (*20*), LIDAR has the potential to reveal many more discoveries in the remaining 83% of the basin that is opaque to other remote sensing approaches.

Here, we report a large number of previously undocumented pre-Columbian earthworks with geometrically patterned enclosures in an Amazon-wide LIDAR dataset covering 0.08% of the basin (*21*). We combine these newly discovered sites with a comprehensive dataset of existing archaeological sites (ring ditches, geoglyphs, ponds, and wells) to model areas likely to harbor as yet undetected earthworks hidden beneath remote forest landscapes. On the basis of our predictive model, we estimate the number of undocumented earthworks and identify domesticated tree species associated with earthwork presence.

Archaeological discoveries beneath the canopy

Scanning 5315 km^2 of LIDAR data originally obtained for estimating aboveground biomass throughout the Amazonian forest (*22*) revealed

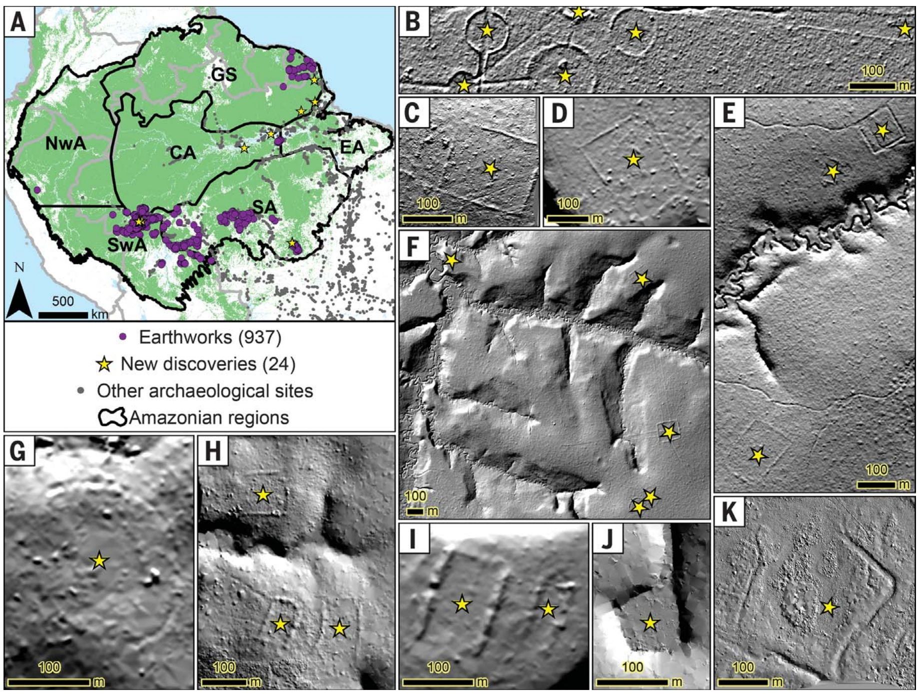


Fig. 1. Geographical distribution of known and newly discovered pre-Columbian geometric earthworks in Amazonia. (**A**) Map of previously reported and newly discovered earthworks (purple circles and yellow stars, respectively) reported in this study across six Amazonian regions: central Amazonia (CA), eastern Amazonia (EA), Guiana Shield (GS), northwestern Amazonia (NwA), southern Amazonia (SA), and southwestern Amazonia (SwA). (**B**) Newly discovered earthworks in SA. (**C** to **F**) Newly discovered earthworks in SwA. (**G** to **I**) Newly discovered earthworks in GS. (**J** and **K**) Newly discovered earthworks in CA. Scale bars, 100 m.

All authors and affiliations are listed at the end of this Research Article.

24 unreported earthworks in southern, southwestern, central, and northern (the Guiana Shield) Amazonia (Fig. 1A) (*21*). We detected a fortified village in southern Amazonia (Fig. 1B), defensive and ceremonial sites in southwestern Amazonia (Fig. 1, C to F), crowned mountains and megalithic structures in the Guiana Shield (Fig. 1, G to I), and riverine sites on floodplains in central Amazonia (Fig. 1, J and K).

In southern Amazonia, we found an ancient plaza town located in the Upper Xingu Basin (Fig. 1B). This region is known to have supported dense populations in the past, distributed throughout plaza villages interconnected by road networks and surrounded by domesticated landscapes with a diverse array of terrestrial and aquatic resources (*10*, *23*). It is also clear that the earthworks in this region extend beyond the sampled area of the 200-m-wide LIDAR transect, restraining their full identification. The layout of these earthworks is similar to that of other fortified villages documented in this region, which supports the idea that these structures were built before European contact (*10*, *15*, *24*).

In southwestern Amazonia, we found a combination of rectangular and circular features, known as geoglyphs, without detectable interconnecting roads occurring on flat terrain close to water bodies (Fig. 1, C to F). Documented defensive and ceremonial earthworks in this region were built around two millennia ago and are dispersed across the well-drained plateaus of the tributaries of the Purus and Madeira rivers (*25*).

In the Guiana Shield, we detected a combination of rectangular and circular features on plateaus near water bodies (Fig. 1, G to I). The region holds different types of earthworks with different usages: permanent settlements within crowned mountains in French Guiana (*26*) and ceremonial sites featuring megalithic structures arranged in circular clusters found along the coast of Amapá, Brazil (*27*).

In the floodplains of central Amazonia, a hotspot of pre-Columbian riverine settlements (*3*, *23*, *28*), we identified two other earthworks (Fig. 1, J and K). We considered these sites to be anthropogenic because of their straight edges, although the geometry of these sites is distinct from that of the earthworks found in upland forests. Constant sedimentary deposition over the centuries, through periodic floods, may have buried smaller features, preserving only the observed structures, which elsewhere have been associated with pre-Columbian fisheries management (*29*).

Modeling basin-wide distribution of earthworks

By extrapolating the density of earthworks observed in our LIDAR data (0.0062 earthworks/km^2) to the extent of Amazonia (6.7 million km^2), we calculated that >41,000 earthworks may occur throughout the forest. However, given that our LIDAR data covered only 0.08% of the total area of Amazonia and that earth-building societies were not evenly distributed across the basin (*15*, *30*), more-rigorous methods were needed to estimate how many other as yet undocumented pre-Columbian earthworks might occur and where. To answer these questions, we used newly developed Bayesian statistical techniques and an inhomogeneous Poisson process (IPP) model (*31*), with an intensity function using intensity covariates and thinned by observability covariates (*32*). Recently, the use of other machine learning techniques such as random forests have become popular for species distribution models (SDMs). There is still some uncertainty about this use (*33*), and the implementation of random forests to IPPs is still not available, but it might be a welcome addition to the toolkit of SDM analysis.

The aforementioned statistical analysis was based on the records of 937 known earthworks complemented by our discoveries (24 earthworks), with three bioclimatic, three edaphic, and three topographic variables as intensity covariates. More than 40 variables were considered in the model (table S1), and the selected ones (nine variables) cover gradients of temperature, precipitation, soil structure and fertility, topography, water-table depth, and distance to water bodies (*21*). Observability

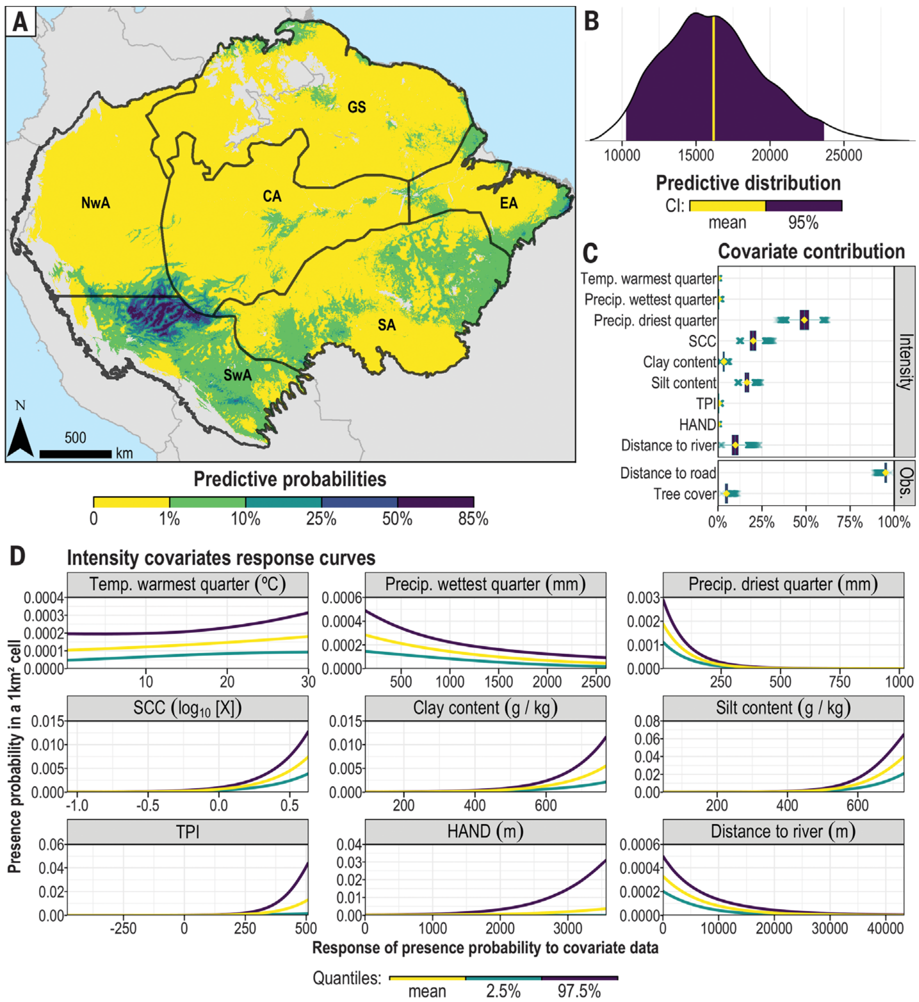


Fig. 2. Probability model of pre-Columbian earthworks across Amazonia. (**A**) Predicted probability of earthwork presence for 1-km^2 cells across six Amazonian regions using an inhomogeneous Poisson process predictive model: central Amazonia (CA), eastern Amazonia (EA), Guiana Shield (GS), northwestern Amazonia (NwA), southern Amazonia (SA), and southwestern Amazonia (SwA). Areas not modeled (NA) are greyed out. (**B**) Predictive probability function for the number of as yet undetected earthworks; the dark area under the curve represents the credibility interval (CI) of the probabilities associated with each number. (**C**) Boxplot of the estimated relative contribution of each covariate; the yellow diamond indicates the mean value. SCC, soil cation concentration; TPI, terrain position index; HAND, height above the nearest drainage. (**D**) Individual predicted probability of earthwork presence against intensity covariates. For projected areas across each Amazonian region on different probability thresholds, see table S2, and for the IPP model on continuous values, see fig. S1.



Fig. 3. Significant relationships between the occurrence and abundance of domesticated tree species and the modeled distribution of earthworks in Amazonia. Point estimates and confidence intervals of species significantly associated with predicted probability of earthwork presence, with an overall significance level of 5%. Positive species are more likely to occur and be abundant where predicted probability of earthwork presence is high, whereas negative species are less likely to occur and be abundant there.

covariates were used to describe the dataset sample preference by indicating the most favorable location for sample acquisition (*32*). The effect of sample selection bias was individually weighted for each sample (*21*).

Our model predicts the number of as yet undiscovered pre-Columbian structures at between 10,272 and 23,648, with 95% probability, giving an average of 16,187 sites (Fig. 2B). These estimates suggest that the earthworks already documented in the Amazon to date account for a mere 4 to 9% of the total, and that 91 to 96% of Amazonian earthworks remain undiscovered.

This predictive model indicated that earthworks are likely concentrated in southwestern Amazonia (Fig. 2A) and corroborated previous studies that found this region to be a hotspot of earth-building societies (*13*, *15*, *34*). In addition, nearly all the highest-probability cells (≥25% predicted probability) occur in a 94,713-km^2 rectangle that overlays a substantial portion of the Brazilian state of Acre. Indeed, southwestern Amazonia contains the earliest plant cultivation and domestication (*9*, *35*), the oldest anthropogenic soils (*35*), low-density urbanism (*19*), and now a much higher density of earthworks. The underlying spatial data distribution may offer valuable information about pre-Columbian practices before European contact (*36*).

Our analysis also suggests that pre-Columbian societies engaged in earthwork construction in all other regions, covering a broader area than previously thought. However, earthworks are heterogeneously distributed across Amazonian regions. Almost 80% of the basin has a 0 to 1% predicted probability of earthwork presence for 1-km^2 cells. These low-probability areas are mostly located in northwestern, northern, and central Amazonia, whereas higher-probability areas (≥25% predicted probability, covering 1.41% of the basin) are located in southwestern Amazonia. Earth-building societies were very common in some parts of the basin, but they may not have occupied all of Amazonia (*6*, *15*, *30*, *37*). Other types of domesticated landscapes, such as Amazonian dark earths, are widespread [see maps in (*37*–*39*)] in regions (e.g., central Amazonia) where the earthworks analyzed in our study (ring ditches, geoglyphs, ponds, and wells) are not commonly found. Given the diversity of pre-Columbian societies and their land-use practices over 12,000 years of ancient Amazonian history, forests were likely modified at varying intensities by different Indigenous populations through time (*7*, *38*).

Forests modified by earth-building societies are more likely to occur in locations with high temperature and low precipitation during the wettest and driest quarters (Fig. 2, C and D). Areas with high soil content of clay and silt and high cation concentrations also show high probabilities of earthwork presence. In addition, earthworks tend to be located on plateaus with deep water tables, yet close to water bodies. This combination of environmental conditions probably facilitated the construction of earthworks by offering periods with less precipitation and higher temperature, and soils with a better texture for earthmoving. In addition, the presence of a drier season facilitates burning, which could help remove the vegetation for building earth structures (*12*), while higher soil cation concentrations could attract settlements for the development of diversified food production systems with plants

managed and domesticated to different degrees (*15*, *30*).

As expected, observability covariates indicate that previously reported earthworks are mostly found near roads, which facilitate field research (Fig. 2C). Tree cover, however, has no effect on the current distribution of earthworks. Thus, new earthworks can still be found even in deforested areas. The use of conventional very-high-resolution remote sensing data, guided by the probability surfaces produced here (Fig. 2A), is likely to reveal more previously undetected earthworks in both closed-canopy and deforested areas of Amazonia. In addition, the rise of machine learning techniques applied to archaeological site detection may lead to rapid discovery of new sites across deforested areas (*40*, *41*).

In forested areas, LIDAR surveys guided by our discoveries (e.g., full coverage of the Fig. 1B site) and the probability surfaces in Fig. 2A are promising tools for discovering new sites. However, very-high-probability areas (≥50% predicted probability) cover 32,120 km^2, for which a complete LIDAR survey would require six times more data than have been collected to date in the Amazon. Thus, other approaches, such as mapping the distribution and abundance of domesticated species associated with earthwork presence, may help locate new sites within the Amazonian forest (*42*, *43*).

Relationships with domesticated species

We analyzed the relationship between the response (occurrence and abundance) of 79 domesticated tree species identified across 1676 forest plots (*6*) and the predicted probability of earthwork presence using generalized linear models to test whether forests with a higher probability of earthwork presence have a higher frequency and abundance of domesticated species (*21*). The occurrence and/or abundance of 35 domesticated species increased with the predicted probability of earthwork presence, while those of 18 species decreased. In total, the occurrence and/or abundance of 53 of the 79 domesticated species showed significant association with the predictive model of earthwork distribution (Fig. 3).

The species whose responses increased the most significantly along with the probability of earthwork occurrence are *Bertholletia excelsa* ($P < 0.001$, $\beta = 1.13$), *Hevea brasiliensis* ($P < 0.001$, $\beta = 0.65$), and *Brosimum alicastrum* ($P < 0.001$, $\beta = 1.36$), on the basis of occurrence data, and *Astrocaryum murumuru* ($P < 0.001$, $\beta = 0.71$), *Attalea phalerata* ($P < 0.001$, $\beta = 1.42$), and *Theobroma cacao* ($P < 0.001$, $\beta = 1.43$), on the basis of abundance data (fig. S2). The species whose responses decreased the most significantly are *Erisma japura* ($P < 0.001$, $\beta = -1.94$), on the basis of occurrence data, and *E. japura* ($P < 0.001$, $\beta = -1.7$) and *Oenocarpus bataua* ($P < 0.001$, $\beta = -0.27$), on the basis of abundance data (fig. S2). Although these highlighted species have multiple uses (*44*), they have mainly been used for their edible fruits and nuts in Amazonia, with the exception of *H. brasiliensis*, which has been used intensively for latex production (data S1). Species that are more frequent and abundant in forests with higher probability of earthwork occurrence were probably favored by a combination of interacting past Indigenous management practices and ecological processes (*6*). These results confirm previous archaeobotanical and ethnobotanical data that have already shown that some species (e.g., *B. excelsa*, *Astrocaryum* spp., and *Attalea* spp.) are more abundant on and near archaeological sites across Amazonia (*8*, *14*, *36*). Species that are less frequent and abundant in areas with a higher probability of earthwork occurrence likely prefer habitats where earthworks are usually not found, such as sandy soils with lower fertility (*7*), or were disfavored by past practices that might have had detrimental effects on some species (*45*).

Social-ecological implications

The massive extent of archaeological sites and widespread human-modified forests across Amazonia is critically important for establishing an accurate understanding of interactions between human societies, Amazonian forests, and Earth's climate (*37*). Considering the widespread extent of locations modified by pre-Columbian management and cultivation practices, Amazonia can be viewed as an ancient social-ecological system, with long-term responses to climate change (*46*), more similar to old secondary forests than pristine climax ecosystems (*10*).

The discovery of earthworks hidden beneath dense forest canopies also indicates that, given sufficient time after these sites became depopulated, forests regenerated over the centuries. It is still unknown, however, the scale of structural and floristic differences between pristine and domesticated forests across Amazonia. The forest reclaimed the land, but this is not the case for the Indigenous societies that managed these forests and waterbodies and that created these large structures. These archaeological legacies can play a role in present-day debates around Indigenous territorial rights. They serve as tangible proof of an ancestor's occupation, way of life, and their relationship with the forest. Today, Indigenous peoples struggle to recognize their right to land originally inhabited by their ancestors, along with the protection of their territories, languages, cultures, and heritages. In addition to protecting the native peoples that remain, the institution of Indigenous lands also collaborates with forest conservation in times of debates on climate change and the search for solutions that minimize impacts on the climate and promote carbon neutrality.

These human-modified landscapes harbor an impressive archaeological heritage. Of the 24 earthworks newly reported in our study, 50% are located in areas with some degree of legal protection. When all 937 known earthworks are considered, however, only 9% are located inside Indigenous lands and protected areas. To date, most pre-Columbian earthworks have been discovered after deforestation. The highest density of known earthworks in Amazonia is, therefore, outside protected areas and mostly located in the region with the highest historical and current rates of deforestation, called the "Arc of Deforestation." Protected areas and Indigenous territories can act as barriers against illegal activities that promote the degradation and destruction of Amazonia's natural and cultural heritage, but their implementation and expansion depend on strong government policies and law enforcement (*47*, *48*).

Ironically, modern-day deforestation is removing the very evidence of pre-Columbian land-use strategies that were able to transform the landscape without causing large-scale deforestation (*13*). Today, Amazonia is experiencing expansion of agriculture and cattle ranching (*49*, *50*), especially where earthworks are concentrated in the southern and southwestern regions, risking the destruction of earthworks and fracturing and hampering the identification of pre-Columbian occupation sites that provide direct evidence of ancient Indigenous territories. Our data on earthwork probability, suitable environmental conditions, and associated domesticated species should narrow the search for Indigenous heritage sites, enhanced by optical and LIDAR sensing to identify, monitor, and help conserve archaeological features. Amazonian forests clearly merit protection not only for their ecological and environmental value but also for their high archaeological, social, and biocultural value, which can teach modern society how to sustainably manage its natural resources.

REFERENCES AND NOTES

1. E. G. Neves *et al*., "Chapter 8: Peoples of the Amazon before European colonization" in *Amazon Assessment Report 2021* (UN Sustainable Development Solutions Network, 2021); https://doi.org/10.55161/LXIT5573.
2. C. L. Erickson, in *Time and Complexity in Historical Ecology: Studies in the Neotropical Lowlands*, W. Balée, C. L. Erickson, Eds., Historical Ecology Series (Columbia Univ. Press, 2006), pp. 235–278.
3. E. G. Neves, in *The Handbook of South American Archaeology*, H. Silverman, W. H. Isbell, Eds. (Springer, 2008), pp. 359–379.
4. D. P. Schaan, *Sacred Geographies of Ancient Amazonia: Historical Ecology of Social Complexity* (Routledge, 2016).
5. C. de P. Moraes, E. G. Neves, in *Encyclopedia of Global Archaeology*, C. Smith, Ed. (Springer International Publishing, Cham, 2020), pp. 3491–3503.
6. C. Levis *et al*., *Science* **355**, 925–931 (2017).
7. C. Levis *et al*., *Front. Ecol. Evol.* **5**, 171 (2018).
8. C. R. Clement, *Econ. Bot.* **53**, 188–202 (1999).
9. U. Lombardo *et al*., *Nature* **581**, 190–193 (2020).
10. M. J. Heckenberger *et al*., *Science* **301**, 1710–1714 (2003).
11. M. J. Heckenberger, J. C. Russell, J. R. Toney, M. J. Schmidt, *Philos. Trans. R. Soc. London Ser. B* **362**, 197–208 (2007).
12. J. F. Carson *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 10497–10502 (2014).

13. J. Watling *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 1868–1873 (2017).
14. C. C. Mann, *Science* **321**, 1148–1152 (2008).
15. J. G. de Souza *et al.*, *Nat. Commun.* **9**, 1125 (2018).
16. J. Iriarte *et al.*, *J. Comput. Appl. Archaeol.* **3**, 151–169 (2020).
17. M. A. Canuto *et al.*, *Science* **361**, eaau0137 (2018).
18. A. F. Chase, D. Z. Chase, C. T. Fisher, S. J. Leisz, J. F. Weishampel, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12916–12921 (2012).
19. H. Prümers, C. J. Betancourt, J. Iriarte, M. Robinson, M. Schaich, *Nature* **606**, 325–328 (2022).
20. C. M. Souza Jr. *et al.*, *Remote Sens.* **12**, 2735 (2020).
21. Materials and methods are available as supplementary materials.
22. G. Tejada, E. B. Görgens, F. D. B. Espírito-Santo, R. Z. Cantinho, J. P. Ometto, *Carbon Balance Manag.* **14**, 11 (2019).
23. M. J. Heckenberger, J. B. Petersen, E. G. Neves, *Lat. Am. Antiq.* **10**, 353–376 (1999).
24. M. J. Heckenberger *et al.*, *Science* **321**, 1214–1217 (2008).
25. S. Saunaluoma, M. Pärssinen, D. Schaan, *J. Field Archaeol.* **43**, 362–379 (2018).
26. G. Odonne, J.-F. Molino, *Les Nouv. l'archéologie* **152**, 11–15 (2018).
27. M. P. Cabral, J. D. M. Saldanha, *Rev. Arqueol. Pública* **3**, 7–13 (2015).
28. P. Stenborg, D. P. Schaan, C. G. Figueiredo, *J. Field Archaeol.* **43**, 44–57 (2018).
29. R. Blatrix *et al.*, *Sci. Rep.* **8**, 5998 (2018).
30. C. H. McMichael, M. W. Palace, M. Golightly, *J. Biogeogr.* **41**, 1733–1745 (2014).
31. N. A. C. Cressie, in *Statistics for Spatial Data*, Wiley Series in Probability and Statistics (John Wiley & Sons, Inc., revised ed., 2015), pp. 575–723.
32. G. A. Moreira, D. Gamerman, *Ann. Appl. Stat.* **16**, 1848–1867 (2022).
33. R. Valavi, J. Elith, J. J. Lahoz-Monfort, G. Guillera-Arroita, *Ecography* **44**, 1731–1742 (2021).
34. S. Y. Maezumi *et al.*, *Nat. Plants* **4**, 540–547 (2018).
35. J. Watling *et al.*, *PLOS ONE* **13**, e0199868 (2018).
36. P. Riris, *J. Anthropol. Archaeol.* **59**, 101177 (2020).
37. C. N. H. McMichael, F. Matthews-Bird, W. Farfan-Rios, K. J. Feeley, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 522–527 (2017).
38. A. M. G. A. WinklerPrins, C. Levis, *Ann. Am. Assoc. Geogr.* **111**, 858–868 (2021).
39. J. Iriarte *et al.*, *Quat. Sci. Rev.* **248**, 106582 (2020).
40. H. A. Orengo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 18240–18250 (2020).
41. A. Bonhage *et al.*, *Archaeol. Prospect.* **28**, 177–186 (2021).
42. M. P. Ferreira, M. Zortea, D. C. Zanotta, Y. E. Shimabukuro, C. R. de Souza Filho, *Remote Sens. Environ.* **179**, 66–78 (2016).
43. M. P. Ferreira *et al.*, *Ecol. Inform.* **63**, 101302 (2021).
44. S. D. Coelho *et al.*, *PLOS ONE* **16**, e0257875 (2021).
45. G. Odonne *et al.*, *Ecology* **100**, e02806 (2019).
46. R. J. W. Brienen *et al.*, *Nature* **519**, 344–348 (2015).
47. K. V. Conceição *et al.*, *Land Use Policy* **108**, 105663 (2021).
48. G. de Oliveira *et al.*, *Forests* **13**, 16 (2021).
49. C. H. L. Silva Junior *et al.*, *Nat. Ecol. Evol.* **5**, 144–145 (2021).
50. G. Mataveli, G. de Oliveira, *Science* **375**, 275–276 (2022).
51. V. Peripato *et al.*, Data from: Over 10,000 Pre-Columbian earthworks are still hidden throughout Amazonia, version 1.0.0, Zenodo (2023); https://doi.org/10.5281/zenodo.7750985.

ACKNOWLEDGMENTS

This paper is the result of the work of hundreds of different scientists and research institutions in the Amazon over the past 80 years. Without their hard work, this analysis would have been impossible. We thank members of the following projects and groups for providing data and support: Sustainable Landscapes Brazil project; Center for Science of the Terrestrial System; TRopical Ecosystems and Environmental Sciences; Amazon Tree Diversity Network; Amazonian Archaeological Sites Network; Pre-Columbian Amazon-Scale Transformations, Biodiversity Research Program in Western Amazon Center for Integrated Studies of Amazonian Biodiversity (PPBio-AmOc/INCT-CENBAM); and the Brazilian Space Agency (AEB). **Funding:** V.P. and C.L. were supported by the Coordination of Superior Level Staff Improvement under Academic Excellence Program (CAPES/PROEX) research grants (1681023, 88887.479608/2020, and 88887.474568/2020); C.L. was supported by National Council for Scientific and Technological Development (CNPQ) grants (159440/2018-1, and 400369/2021-4); D.Gam. was supported by a CNPQ grant (304742/2018-0); A.B.J. was supported by the European Research Council (ERC) under a Consolidator Grant (FP7-771056-LICCI); J.P.H. B.O. was supported by a São Paulo Research Foundation (FAPESP) grant (2017/22269-2) and Airborne LASER Scanning data acquisition by the Amazon Fund grant (14.2.0929.1); H.L.G.C. was supported by a FAPESP grant (18/14423-4); H.t.S., V.H.F.G., and R.S. were supported by a PVE-MEC/MCTI/CAPES/CNPq/FAPs grant (407232/2013-3); J.G.d.S., J.I., and M.Rob. were supported by Horizon 2020 grants (ERC Cog 616179 and ERC PoC_777845); A.M.d.S. was supported by a CAPES grant (88887.607664/2021); D.S., J.-F.M., J.E., P.P., and J.C. were supported by an ANR grant (CEBA 10-LABX-25-01); H.L.d.Q. and J.L.L.M. were supported by MCT/CNPq/CT-INFRA/GEOMA grants (550373/2010-1 and 457515/2012-0); J.L.L.M. was supported by a CAPES/PDSE grant (88881.135761/2016-01) and a CAPES/Fapespa grant (1530801); E.M.V. was supported by a CNPq grant (308040/2017-1); B.M.F. was supported by a FAPESP grant (2016/25086-3); B.S.M., B.H.M.-J., and O.L.P. were supported by a CNPq/CAPES/FAPS/BC-Newton grant (441244/2016-5), a FAPEMAT grant (0589267/2016), and a Royal Society GCRF International Collaboration Award (ICA\R1\180100); T.W.H. was supported by an NSF/DEB grant (1556338); L.E.O.C.A. was supported by a CNPQ/PQ grant (314416/2020-0). Floristic identification in plots in the RAINFOR forest monitoring network and plot data management by ForestPlots.net have been supported by several Natural Environment Research Council grants to O.L.P. and colleagues (NE/B503384/1, NE/D01025X/1, NE/I02982X/1, NE/F005806/1, NE/D005590/1, NE/I028122/1, and NE/S011811/1) and the Gordon and Betty Moore Foundation. **Author contributions:** Conceptualization: V.P., C.L., and L.E.O.C.A. LIDAR raw data processing: V.P. Earthworks investigation: V.P., J.G.d.S., J.I., M.Rob., and L.E.O.A.C. Modeling: G.A.M. and D.Gam. Domestication investigation: V.P. and C.L. Writing – original draft: V.P., C.L., and L.E.O.A.C. Writing – review & editing: V.P., C.L., G.A.M., D.Gam., N.C.A.P., and L.E.O.A.C. All of the other authors contributed data, discussed further analyses, and commented on various versions of the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Data from publicly available sources are cited in the supplementary materials. Other data and computer codes used in the analysis are publicly available in Zenodo (*51*).

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Authors:
Vinicius Peripato[1]*†, Carolina Levis[2]†, Guido A. Moreira[3], Dani Gamerman[4], Hans ter Steege[5], Nigel C. A. Pitman[6], Jonas G. de Souza[7], José Iriarte[8], Mark Robinson[8], André Braga Junqueira[9], Thiago B. Trindade[10], Fernando O. de Almeida[11], Claide de Paula Moraes[12], Umberto Lombardo[13], Eduardo K. Tamanaha[14], Shira Y. Maezumi[15], Jean P. H. B. Ometto[1], José R. G. Braga[1], Wesley A. Campanharo[1], Henrique L. G. Cassol[1], Philipe R. Leal[1], Mauro L. R. de Assis[1], Adriana M. da Silva[16], Oliver L. Phillips[17], Flávia R. C. Costa[18], Bernardo Monteiro Flores[2], Bruce Hoffman[19], Terry W. Henkel[20], Maria Natalia Umaña[21], William E. Magnusson[18], Elvis H. Valderrama Sandoval[22,23], Jos Barlow[24], William Milliken[25], Maria Aparecida Lopes[26], Marcelo Fragomeni Simon[27], Tinde R. van Andel[5,28], Susan G. W. Laurance[29], William F. Laurance[29], Armando Torres-Lezama[30], Rafael L. Assis[31], Jean-François Molino[32], Mickaël Mestre[33], Michelle Hamblin[34], Luiz de Souza Coelho[35], Diogenes de Andrade Lima Filho[35], Florian Wittmann[36,37], Rafael P. Salomão[38,39], Iêda Leão Amaral[35], Juan Ernesto Guevara[40,41], Francisca Dionízia de Almeida Matos[35], Carolina V. Castilho[42], Marcelo de Jesus Veiga Carim[43], Dairon Cárdenas López[44]‡, Daniel Sabatier[32], Mariana Victória Irume[35], Maria Pires Martins[35], José Renan da Silva Guimarães[45], Olaf S. Bánki[5], Maria Teresa Fernandez Piedade[37], José Ferreira Ramos[35], Bruno Garcia Luize[46], Evlyn Márcia Moraes de Leão Novo[1], Percy Núñez Vargas[47], Thiago Sanna Freire Silva[48], Eduardo Martins Venticinque[49], Angelo Gilberto Manzatto[50], Neidiane Farias Costa Reis[51], John Terborgh[52,29], Katia Regina Casula[51], Layon O. Demarchi[37], Euridice N. Honorio Coronado[53,54], Abel Monteagudo Mendoza[47,55], Juan Carlos Montero[56,35], Jochen Schöngart[37], Ted R. Feldpausch[57,17], Adriano Costa Quaresma[36,37], Gerardo A. Aymard C.[58], Chris Baraloto[59], Nicolás Castaño Arboleda[44], Julien Engel[32,59], Pascal Petronelli[60], Charles Eugene Zartman[35], Timothy J. Killeen[61], Beatriz S. Marimon[62], Ben Hur Marimon-Junior[62], Juliana Schietti[35], Thaiane R. Sousa[63], Rodolfo Vasquez[55], Lorena M. Rincón[35], Erika Berenguer[64,24], Joice Ferreira[65], Bonifacio Mostacedo[66], Dário Dantas do Amaral[39], Hernán Castellanos[67], Marcelo Brilhante de Medeiros[27], Ana Andrade[68], José Luís Camargo[68], Emanuelle de Sousa Farias[69,70], José Leonardo Lima Magalhães[71,65], Henrique Eduardo Mendonça Nascimento[35], Helder Lima de Queiroz[72], Roel Brienen[17], Juan David Cardenas Revilla[35], Pablo R. Stevenson[73], Alejandro Araujo-Murakami[74], Bruno Barçante Ladvocat Cintra[75], Yuri Oliveira Feitosa[76], Flávia Rodrigues Barbosa[77], Rainiellen de Sá Carpanedo[77], Joost F. Duivenvoorden[78], Janaína da Costa de Noronha[77], Domingos de Jesus Rodrigues[77], Hugo F. Mogollón[79], Leandro Valle Ferreira[39], John Ethan Householder[36], José Rafael Lozada[80], James A. Comiskey[81,82], Freddie C. Draper[8382], José Julio de Toledo[84], Gabriel Damasco[85], Nállarett Dávila[46]§, Roosevelt García-Villacorta[86,87], Aline Lopes[88], Fernando Cornejo Valverde[89], Alfonso Alonso[82], Francisco Dallmeier[82], Vitor H. F. Gomes[90,91], Eliana M. Jimenez[92], David Neill[93], Maria Cristina Peñuela Mora[94], Daniel P. P. de Aguiar[95,96], Luzmila Arroyo[74], Fernanda Antunes Carvalho[18,97], Fernanda Coelho de Souza[18,17], Kenneth J. Feeley[98,99], Rogerio Gribel[35], Marcelo Petratti Pansonato[35,100], Marcos Ríos Paredes[101], Izaias Brasil da Silva[102], Maria Julia Ferreira[103], Paul V. A. Fine[104], Émile Fonty[105,32], Marcelino Carneiro Guedes[106], Juan Carlos Licona[56], Toby Pennington[57,107], Carlos A. Peres[108], Boris Eduardo Villa Zegarra[109], Germaine Alexander Parada[74], Guido Pardo Molina[110], Vincent Antoine Vos[110], Carlos Cerón[111], Paul Maas[5], Marcos Silveira[112], Juliana Stropp[113], Raquel Thomas[114], Tim R. Baker[17], Doug Daly[115], Isau Huamantupa-Chuquimaco[116], Ima Célia Guimarães Vieira[39], Bianca Weiss Albuquerque[37], Alfredo Fuentes[117,118], Bente Klitgaard[119], José Luis Marcelo-Peña[120], Miles R. Silman[121], J. Sebastián Tello[118], Corine Vriesendorp[6], Jerome Chave[122], Anthony Di Fiore[123,124], Renato Richard Hilário[84], Juan Fernando Phillips[125], Gonzalo Rivas-Torres[124,126], Patricio von Hildebrand[127], Luciana de Oliveira Pereira[57], Edelcilio Marques Barbosa[35], Luiz Carlos de Matos Bonates[35], Hilda Paulette Dávila Doza[101], Ricardo Zárate Gómez[128], George Pepe Gallardo Gonzales[101], Therany Gonzales[129], Yadvinder Malhi[130], Ires Paula de Andrade Miranda[35], Linder Felipe Mozombite Pinto[101], Adriana Prieto[131], Agustín Rudas[131], Ademir R. Ruschel[65], Natalino Silva[132], César I. A. Vela[133], Egleé L. Zent[134], Stanford Zent[134], Angela Cano[73,135], Yrma Andreina Carrero Márquez[136], Diego F. Correa[73,137], Janaina Barbosa Pedrosa Costa[106], David Galbraith[17], Milena Holmgren[138], Michelle Kalamandeen[139], Guilherme Lobo[37], Marcelo Trindade Nascimento[140], Alexandre A. Oliveira[100], Hirma Ramirez-Angulo[30], Maira Rocha[37], Veridiana Vizoni Scudeller[141], Rodrigo Sierra[142], Milton Tirado[142], Geertje van der Heijden[143], Emilio Vilanova Torre[30,144], Manuel Augusto Ahuite Reategui[145], Cláudia Baider[146,100], Henrik Balslev[147], Sasha Cárdenas[73], Luisa Fernanda Casas[73], William Farfan-Rios[47,118,148], Cid Ferreira[35], Reynaldo Linares-Palomino[82], Casimiro Mendoza[149,150], Italo Mesones[104], Ligia Estela Urrego Giraldo[151], Daniel Villarroel[74,152], Roderick Zagt[153], Miguel N. Alexiades[154], Edmar Almeida de Oliveira[62], Karina Garcia-Cabrera[121], Lionel Hernandez[67], Walter Palacios Cuenca[155], Susamar Pansini[51], Daniela Pauletto[156], Fredy Ramirez Arevalo[23], Adeilza Felipe Sampaio[51], Luis Valenzuela Gamarra[55], Luiz E. O. C. Aragão[1,57]*

[1]Division of Earth Observation and Geoinformatics, General Coordination of Earth Sciences, National Institute for Space Research (INPE), São José dos Campos, SP, Brazil. [2]Postgraduate Program in Ecology, Federal University of Santa Catarina (UFSC), Florianópolis, SC, Brazil. [3]Centre of Molecular and Environmental Biology, Universidade do Minho, Braga, Portugal. [4]Departamento de Métodos Estatísticos, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brazil. [5]Naturalis Biodiversity Center, Leiden, Netherlands. [6]Quantitative Biodiversity Dynamics, Utrecht University, Utrecht, Netherlands. [7]Science and Education, The Field Museum, Chicago, IL, USA. [8]Department of Humanities, Universitat Pompeu Fabra, Barcelona, Spain. [9]Department of Archaeology, College of Humanities, University of Exeter, Exeter, UK. [10]Institut de Ciència i Tecnologia Ambientals, Universitat Autònoma de Barcelona, Barcelona, Spain. [11]Instituto do Património Histórico e Artístico Nacional (IPHAN), Centro Nacional de Arqueologia (CNA), Brasília, DF, Brazil. [12]Departamento de Arqueologia, Universidade Federal de Sergipe (UFS), Laranjeiras, SE, Brazil. [13]Programa de Antropologia e Arqueologia, Universidade Federal do Oeste do Pará (UFOPA), Santarém, PA, Brazil. [14]Geographisches Institut, University of Bern, Bern, Switzerland. [15]Instituto de Desenvolvimento Sustentável Mamirauá, Tefé, AM, Brazil. [16]Department of Archaeology, Max Planck Institute of Geoanthropology, Jena, Germany. [17]Postgraduate Program in Geography, Institute of Geography, Federal University of Uberlândia (UFU), Uberlândia, MG, Brazil. [18]School of Geography, University of Leeds, Leeds, UK. [19]Coordenação de Pesquisas em Ecologia, Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus, AM, Brazil. [20]Amazon Conservation Team, Arlington, VA, USA. [21]Department of Biological Sciences, Humboldt State University, Arcata, CA, USA. [22]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA. [23]Department of Biology, University of Missouri, St. Louis, MO, USA. [24]Facultad de Biologia, Universidad Nacional de la Amazonia Peruana, Iquitos, Loreto, Peru. [25]Lancaster Environment Centre, Lancaster

University, Lancaster, Lancashire, UK. [26]Department for Ecosystem Stewardship, Royal Botanic Gardens, Richmond, Surrey, UK. [27]Instituto de Ciências Biológicas, Universidade Federal do Pará (UFPA), Belém, PA, Brazil. [28]Embrapa Recursos Genéticos e Biotecnologia, Parque Estação Biológica, Prédio da Botânica e Ecologia, Brasilia, DF, Brazil. [29]Biosystematics Group, Wageningen University, Wageningen, Netherlands. [30]Centre for Tropical Environmental and Sustainability Science and College of Science and Engineering, James Cook University, Cairns, QLD, Australia. [31]Instituto de Investigaciones para el Desarrollo Forestal (INDEFOR), Universidad de los Andes, Conjunto Forestal, Mérida, Mérida, Venezuela. [32]Biodiversity and Ecosystem Services, Instituto Tecnológico Vale, Belém, PA, Brazil. [33]AMAP, IRD, Cirad, CNRS, INRAE, Université de Montpellier, Montpellier, France. [34]Institut National de Recherches Archéologiques Préventives, Bègles, France. [35]Direction des Affaires Culturelles (DAC Guyane), Cayenne, French Guiana. [36]Coordenação de Biodiversidade, Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus, AM, Brazil. [37]Wetland Department, Institute of Geography and Geoecology, Karlsruhe Institute of Technology (KIT), Rastatt, Germany. [38]Ecology, Monitoring and Sustainable Use of Wetlands (MAUA), Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus, AM, Brazil. [39]Programa de Pós-Graduação em Ciências Biológicas e Botânica Tropical, Universidade Federal Rural da Amazônia (UFRA), Belém, PA, Brazil. [40]Coordenação de Botânica, Museu Paraense Emílio Goeldi, Belém, PA, Brazil. [41]Grupo de Investigación en Biodiversidad, Medio Ambiente y Salud (BIOMAS), Universidad de las Américas, Campus Queri, Quito, Ecuador. [42]Centro de Pesquisa Agroflorestal de Roraima, Embrapa Roraima, Boa Vista, RR, Brazil. [43]Departamento de Botânica, Instituto de Pesquisas Científicas e Tecnológicas do Amapá (IEPA), Macapá, AP, Brazil. [44]Herbario Amazónico Colombiano, Instituto Amazónico de Investigaciones Científicas (SINCHI), Bogotá, DC, Colombia. [45]Amcel Amapá Florestal e Celulose S.A, Santana, AP, Brazil. [46]Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Campinas, SP, Brazil. [47]Herbario Vargas, Universidad Nacional de San Antonio Abad del Cusco (UNSAAC), Cusco, Cusco, Peru. [48]Biological and Environmental Sciences, University of Stirling, Stirling, UK. [49]Departamento de Ecologia, Centro de Biociências, Universidade Federal do Rio Grande do Norte (UFRN), Natal, RN, Brazil. [50]Departamento de Biologia, Universidade Federal de Rondônia (UNIR), Porto Velho, RO, Brazil. [51]Programa de Pós-Graduação em Biodiversidade e Biotecnologia, Universidade Federal de Rondônia (UNIR), Porto Velho, RO, Brazil. [52]Department of Biology and Florida Museum of Natural History, University of Florida, Gainesville, FL, USA. [53]Instituto de Investigaciones de la Amazonía Peruana (IIAP), Iquitos, Loreto, Peru. [54]School of Geography and Sustainable Development, University of St Andrews, St Andrews, UK. [55]Jardín Botánico de Missouri, Oxapampa, Pasco, Peru. [56]Instituto Boliviano de Investigacion Forestal, Santa Cruz, Santa Cruz, Bolivia. [57]Geography, College of Life and Environmental Sciences, University of Exeter, Exeter, UK. [58]Programa de Ciencias del Agro y el Mar, Herbario Universitario (PORT), UNELLEZ-Guanare, Guanare, Portuguesa, Venezuela. [59]International Center for Tropical Botany (ICTB), Department of Biological Sciences, Florida International University, Miami, FL, USA. [60]Paracou research station, UMR EcoFoG Université de Guyane, Campus agronomique, Kourou Cedex, French Guiana. [61]Agteca-Amazonica, Santa Cruz, Bolivia. [62]Programa de Pós-Graduação em Ecologia e Conservação, Universidade do Estado de Mato Grosso (UNEMAT), Nova Xavantina, MT, Brazil. [63]Programa de Pós-Graduação em Ecologia, Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus, AM, Brazil. [64]Environmental Change Institute, University of Oxford, Oxford, Oxfordshire, UK. [65]Empresa Brasileira de Pesquisa Agropecuária, Embrapa Amazônia Oriental, Belém, PA, Brazil. [66]Facultad de Ciencias Agrícolas, Universidad Autónoma Gabriel René Moreno, Santa Cruz, Santa Cruz, Bolivia. [67]Centro de Investigaciones Ecológicas de Guayana, Universidad Nacional Experimental de Guayana, Puerto Ordaz, Bolivar, Venezuela. [68]Projeto Dinâmica Biológica de Fragmentos Florestais, Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus, AM, Brazil. [69]Laboratório de Ecologia de Doenças Transmissíveis da Amazônia (EDTA), Instituto Leônidas e Maria Deane (Fiocruz Amazônia), Manaus, AM, Brazil. [70]Programa de Pós-graduação em Biodiversidade e Saúde, Instituto Oswaldo Cruz (Fiocruz), Rio de Janeiro, RJ, Brazil. [71]Programa de Pós-Graduação em Ecologia, Universidade Federal do Pará (UFPA), Belém, PA, Brazil. [72]Diretoria Técnico-Científica, Instituto de Desenvolvimento Sustentável Mamirauá, Tefé, AM, Brazil. [73]Laboratorio de Ecología de Bosques Tropicales y Primatología, Universidad de los Andes, Bogotá, DC, Colombia. [74]Museo de Historia Natural Noel Kempff Mercado, Universidad Autónoma Gabriel Rene Moreno, Santa Cruz, Santa Cruz, Bolivia. [75]Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo (USP), São Paulo, SP, Brazil. [76]Programa de Pós-Graduação em Botânica, Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus, AM, Brazil. [77]Institute of Natural, Human, and Social Sciences (ICNHS), Federal University of Mato Grosso (UFMT), Sinop, MT, Brazil. [78]Institute of Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, Netherlands. [79]Endangered Species Coalition, Silver Spring, MD, USA. [80]Facultad de Ciencias Forestales y Ambientales, Instituto de Investigaciones para el Desarrollo Forestal, Universidad de los Andes, Mérida, Mérida, Venezuela. [81]Inventory and Monitoring Program, National Park Service, Fredericksburg, VA, USA. [82]Center for Conservation and Sustainability, Smithsonian Conservation Biology Institute, Washington, DC, USA. [83]Department of Geography and Planning, University of Liverpool, Liverpool, UK. [84]Departamento de Meio Ambiente e Desenvolvimento, Universidade Federal do Amapá (UNIFAP), Macapá, AP, Brazil. [85]Gothenburg Global Biodiversity Centre, University of Gothenburg, Gothenburg, Sweden. [86]Programa Restauración de Ecosistemas (PRE), Centro de Innovación Científica Amazónica (CINCIA), Tambopata, Madre de Dios, Peru. [87]Peruvian Center for Biodiversity and Conservation (PCBC), Iquitos, Loreto, Peru. [88]Department of Ecology, Institute of Biological Sciences, University of Brasilia (UNB), Brasilia, DF, Brazil. [89]Andes to Amazon Biodiversity Program, Madre de Dios, Madre de Dios, Peru. [90]Escola de Negócios Tecnologia e Inovação, Centro Universitário do Pará, Belém, PA, Brazil. [91]Environmental Science Program, Geosciences Department, Universidade Federal do Pará (UFPA), Belém, PA, Brazil. [92]Grupo de Ecología y Conservación de Fauna y Flora Silvestre, Instituto Amazónico de Investigaciones Imani, Universidad Nacional de Colombia sede Amazonia, Leticia, Amazonas, Colombia. [93]Universidad Estatal Amazónica, Puyo, Pastaza, Ecuador. [94]Universidad Regional Amazónica IKIAM, Tena, Napo, Ecuador. [95]Procuradoria-Geral de Justiça, Ministério Público do Estado do Amazonas, Manaus, AM, Brazil. [96]Coordenação de Dinâmica Ambiental, Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus, AM, Brazil. [97]Departamento de Genética, Instituto de Ciências Biológicas, Ecologia e Evolução, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brazil. [98]Department of Biology, University of Miami, Coral Gables, FL, USA. [99]Fairchild Tropical Botanic Garden, Coral Gables, FL, USA. [100]Departamento de Ecologia, Instituto de Biociências, Universidade de São Paulo (USP), São Paulo, SP, Brazil. [101]Servicios de Biodiversidad EIRL, Iquitos, Loreto, Peru. [102]Postgraduate Program in Biodiversity and Biotechnology Bionorte, Federal University of Acre (UFAC), Rio Branco, AC, Brazil. [103]Postgraduate Program in Ethnobiology and Nature Conservation, Federal Rural University of Pernambuco (UFRPE), Pernambuco, PB, Brazil. [104]Department of Integrative Biology, University of California, Berkeley, CA, USA. [105]Direction régionale de la Guyane, Office national des forêts, Cayenne, French Guiana. [106]Empresa Brasileira de Pesquisa Agropecuária, Embrapa Amapá, Macapá, AP, Brazil. [107]Tropical Diversity Section, Royal Botanic Garden Edinburgh, Edinburgh, Scotland, UK. [108]School of Environmental Sciences, University of East Anglia, Norwich, UK. [109]Direccíon de Evaluación Forestal y de Fauna Silvestre, Magdalena del Mar, Lima, Peru. [110]Instituto de Investigaciones Forestales de la Amazonía, Universidad Autónoma del Beni José Ballivián, Campus Universitario Final, Riberalta, Beni, Bolivia. [111]Escuela de Biología Herbario Alfredo Paredes, Universidad Central, Quito, Pichincha, Ecuador. [112]Centro de Ciências Biológicas e da Natureza, Universidade Federal do Acre (UFAC), Rio Branco, AC, Brazil. [113]Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain. [114]Iwokrama International Centre for Rain Forest Conservation and Development, Georgetown, Guyana. [115]New York Botanical Garden, Bronx, New York, NY, USA. [116]Herbario HAG, Universidad Nacional Amazónica de Madre de Dios (UNAMAD), Puerto Maldonado, Madre de Dios, Peru. [117]Herbario Nacional de Bolivia, Universitario UMSA, La Paz, La Paz, Bolivia. [118]Center for Conservation and Sustainable Development, Missouri Botanical Garden, St. Louis, MO, USA. [119]Department for Accelerated Taxonomy, Royal Botanic Gardens, Richmond, Surrey, UK. [120]Departamento Académico de Ingenieria Forestal y Ambiental, Universidad Nacional de Jaén, Jaén, Cajamarca, Peru. [121]Biology Department and Center for Energy, Environment and Sustainability, Wake Forest University, Winston Salem, NC, USA. [122]Laboratoire Evolution et Diversité Biologique, Université Paul Sabatier CNRS UMR 5174 EDB, Toulouse, France. [123]Department of Anthropology, University of Texas at Austin, Austin, TX, USA. [124]Estación de Biodiversidad Tiputini, Colegio de Ciencias Biológicas y Ambientales, Universidad San Francisco de Quito-USFQ, Quito, Pichincha, Ecuador. [125]Fundación Puerto Rastrojo, Bogotá, DC, Colombia. [126]Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, FL, USA. [127]Fundación Estación de Biología, Bogotá, DC, Colombia. [128]PROTERRA, Instituto de Investigaciones de la Amazonía Peruana (IIAP), Iquitos, Loreto, Peru. [129]ACEER Foundation, Puerto Maldonado, Madre de Dios, Peru. [130]Environmental Change Institute, Oxford University Centre for the Environment, Oxford, England, UK. [131]Instituto de Ciencias Naturales, Universidad Nacional de Colombia, Bogotá, DC, Colombia. [132]Instituto de Ciência Agrárias, Universidade Federal Rural da Amazônia (UFRA), Belém, PA, Brazil. [133]Escuela Profesional de Ingeniería Forestal, Universidad Nacional de San Antonio Abad del Cusco, Puerto Maldonado, Madre de Dios, Peru. [134]Laboratory of Human Ecology, Instituto Venezolano de Investigaciones Científicas (IVIC), Caracas, DC, Venezuela. [135]Cambridge University Botanic Garden, Cambridge University, Cambridge, UK. [136]Programa de Maestria de Manejo de Bosques, Universidad de los Andes, Mérida, Mérida, Venezuela. [137]Centre for Biodiversity and Conservation Science (CBCS), The University of Queensland, Brisbane, QLD, Australia. [138]Resource Ecology Group, Wageningen University & Research, Wageningen, Gelderland, Netherlands. [139]School of Earth, Environment and Society, McMaster University, Hamilton, Ontario, Canada. [140]Laboratório de Ciências Ambientais, Universidade Estadual do Norte Fluminense (UENF), Campos dos Goytacazes, RJ, Brazil. [141]Departamento de Biologia, Instituto de Ciências Biológicas (ICB), Universidade Federal do Amazonas (UFAM), Manaus, AM, Brazil. [142]GeoIS, Quito, Pichincha, Ecuador. [143]Faculty of Social Sciences, University of Nottingham, University Park, Nottingham, UK. [144]Wildlife Conservation Society (WCS), New York, NY, USA. [145]Medio Ambiente, PLUSPRETOL, Iquitos, Loreto, Peru. [146]The Mauritius Herbarium, Agricultural Services, Ministry of Agro-Industry and Food Security, Reduit, Mauritius. [147]Department of Biology, Aarhus University, Aarhus C, Aarhus, Denmark. [148]Living Earth Collaborative, Washington University in St. Louis, St. Louis, MO, USA. [149]Escuela de Ciencias Forestales (ESFOR), Universidad Mayor de San Simon (UMSS), Sacta, Cochabamba, Bolivia. [150]FOMABO, Manejo Forestal en las Tierras Tropicales de Bolivia, Sacta, Cochabamba, Bolivia. [151]Departamento de Ciencias Forestales, Universidad Nacional de Colombia, Medellín, Antioquia, Colombia. [152]Fundación Amigos de la Naturaleza (FAN), Santa Cruz, Santa Cruz, Bolivia. [153]Tropenbos International, Ede, Netherlands. [154]School of Anthropology and Conservation, University of Kent, Canterbury, Kent, UK. [155]Herbario Nacional del Ecuador, Universidad Técnica del Norte, Quito, Pichincha, Ecuador. [156]Instituto de Biodiversidade e Florestas, Universidade Federal do Oeste do Pará (FOPROP), Campus Tapajós, Santarém, PA, Brazil.

***Corresponding author. Email: vinicius.peripato@gmail.com (V.P.); luiz.aragao@inpe.br (L.E.O.C.A.)**
†These authors contributed equally to this work.
‡Deceased.
§Deceased.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade2541
Materials and Methods
Figs. S1 to S19
Tables S1 to S5
References (*52–80*)
MDAR Reproducibility Checklist
Data S1 and S2

Submitted 16 September 2022; accepted 31 August 2023
10.1126/science.ade2541

BIOCHEMISTRY

Structure of a ribonucleotide reductase R2 protein radical

Hugo Lebrette[1,2]*†, Vivek Srinivas[1]†, Juliane John[1], Oskar Aurelius[1,3], Rohit Kumar[1], Daniel Lundin[1], Aaron S. Brewster[4], Asmit Bhowmick[4], Abhishek Sirohiwal[1], In-Sik Kim[4], Sheraz Gul[4], Cindy Pham[4], Kyle D. Sutherlin[4], Philipp Simon[4], Agata Butryn[5,6]‡, Pierre Aller[5,6], Allen M. Orville[5,6], Franklin D. Fuller[7], Roberto Alonso-Mori[7], Alexander Batyuk[7], Nicholas K. Sauter[4], Vittal K. Yachandra[4], Junko Yano[4], Ville R. I. Kaila[1], Britt-Marie Sjöberg[1], Jan Kern[4]*, Katarina Roos[8]§, Martin Högbom[1]*

Aerobic ribonucleotide reductases (RNRs) initiate synthesis of DNA building blocks by generating a free radical within the R2 subunit; the radical is subsequently shuttled to the catalytic R1 subunit through proton-coupled electron transfer (PCET). We present a high-resolution room temperature structure of the class Ie R2 protein radical captured by x-ray free electron laser serial femtosecond crystallography. The structure reveals conformational reorganization to shield the radical and connect it to the translocation path, with structural changes propagating to the surface where the protein interacts with the catalytic R1 subunit. Restructuring of the hydrogen bond network, including a notably short O–O interaction of 2.41 angstroms, likely tunes and gates the radical during PCET. These structural results help explain radical handling and mobilization in RNR and have general implications for radical transfer in proteins.

Uncontrolled or unmitigated free radicals can cause damage to cells but radicals are also essential to numerous metabolic pathways and enzyme-mediated chemistry (*1*, *2*). Ribonucleotide reductase (RNR) is an archetypal radical enzyme, and the tyrosyl radical (Y•) in the R2 subunit from *Escherichia coli* was the first stable protein radical to be observed, 50 years ago (*3*). RNR provides the only pathway for de novo synthesis of deoxyribonucleotides and represents a drug target for both cancer and infectious diseases (*4*, *5*). Aerobic RNR (class I) depends on the ferritin-like R2 subunit to generate a catalytic radical in an oxygen-dependent reaction; the radical must be transferred back and forth with the catalytic R1 subunit, which performs the ribonucleotide reduction (*6*). Radical translocation between the subunits proceeds through reversible long-range proton-coupled electron transfer (PCET) in a transient R1-R2 complex (*7*). Radical transfer initiation involves redox-induced structural changes in R2 (*8*, *9*), conformational gating (*10*), short-range proton transfer (*11*) coupled to long-distance electron transfer (*12*), and regulation by R1 (*7*). Recently, the structure of an R1-R2 holocomplex was determined by cryo–electron microscopy (cryo-EM) (*7*) providing a picture of the long-range radical transfer pathway. However, atomic resolution snapshots of the radical state and the conformational gating taking place at the R2 active site remain unresolved.

Most R2 proteins harbor a conserved tyrosine residue oxidized to Y• by an oxygen-activated metal center in the catalytically active state. In a recently discovered active metal-free R2 subclass, denoted R2e, this tyrosine residue is post translationally *meta*-hydroxylated to a 3,4-dihydroxyphenylalanine (DOPA) which serves as the radical-harboring residue (*13*, *14*). Oxygen activation of R2e and metal-containing R2 proteins display analogous pathways consisting of a nonactivated state, a catalytically active radical state, and a radical-lost ground state (Fig. 1A).

From an experimental point of view, R2e from *Mesoplasma florum* (*Mf*R2) represents an attractive model to study active radical states in RNRs. Its radical state is metal-independent which simplifies sample preparation and excludes partial occupancy or mismetalation of the metal site often encountered with metalloenzymes in vitro (*15*). In addition, in absence of protein R1 its radical state is stable with no decay observed after more than six hours at 25°C (*13*). However, as with any radical state, it is expected to be highly sensitive to photoreduction as x-ray radiation damage generates free radicals that spread through protein crystals (*16*). This obstacle renders the structural characterization of a protein radical state using standard x-ray crystallography methods unfeasible (*17*, *18*). Though many crystal structures of iron- and manganese-containing R2 proteins from different organisms have been obtained, they describe either the nonactivated state, the radical-lost ground state (often referred to as the "met" state) or partially reduced states. Crystal structures of R2e have been solved from two different organisms (*13*, *14*), likewise showing signs of x-ray–induced photoreduction.

Discrepancies between R2 crystal structures and spectroscopic data of radical states have been observed, leading to contradictory theoretical models regarding the conformation of Y• and its environment (*8*, *19*–*21*). In the present study, by rapid protein production, microcrystallization and x-ray free-electron laser (XFEL) serial femtosecond crystallography, using the diffraction-before-destruction principle (*22*), we have determined the atomic structure of *Mf*R2 in the active radical state at room temperature. Compared with the structure of the radical-lost ground state, also presented here, the radical state structure reveals a notably short hydrogen bond and a critical rearrangement of conserved residues upon acquisition of the radical. Based on these two distinct states, we propose a mechanism for structural recognition of the radical state and a model for redox-coupled conformational gating as a prologue to the radical transfer. This mechanism defines central aspects of the PCET process and may be conserved in aerobic RNRs.

Structure of the radical-lost ground state of *Mf*R2

In the catalytically active radical state, the radical-harboring *meta*-hydroxylated tyrosyl (from here denoted $_{DOPA}$Y126•) in *Mf*R2 exhibits a characteristic absorbance peak at 383 nm and colors the protein blue (*13*). The catalytic radical can be chemically quenched by hydroxyurea, a known RNR radical scavenger used for decades as an antitumor drug (*23*) proposed to inactivate R2 through PCET (*24*). Importantly, hydroxyurea causes a reversible inactivation of *Mf*R2 as the enzyme can recover activity upon reoxidation by NrdI (*13*). Thus, the protein is not permanently inactivated or damaged but resides in a radical-lost ground state (Fig. 1A).

To ensure an accurate depiction of the radical-lost ground state of *Mf*R2, we solved the crystal structure at 1.35 Å resolution of the protein chemically quenched by incubation with hydroxyurea before crystallization. This treatment abolished the 383-nm absorbance peak and rendered the protein colorless (*13*). The electron density map clearly shows the post translational modification in the *meta* position of $_{DOPA}$Y126 (Fig. 1B). The two oxygen-containing functional groups of $_{DOPA}$Y126 in the *para* and *meta* positions (denoted *para*-O and *meta*-O, respectively) form hydrogen bonds (H-bond) with D88. Notably, the *meta*-O is involved in an unusually short interaction (Fig. 1B). Using END/RAPID error analysis (*25*), the O–O distance between $_{DOPA}$Y126 and D88 was

[1]Department of Biochemistry and Biophysics, Stockholm University, Arrhenius Laboratories for Natural Sciences, Stockholm, Sweden. [2]Laboratoire de Microbiologie et Génétique Moléculaires, Centre de Biologie Intégrative, CNRS, Université Toulouse III, Toulouse, France. [3]MAX IV Laboratory, Lund University, Lund, Sweden. [4]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [5]Diamond Light Source Ltd, Harwell Science and Innovation Campus, Didcot, UK. [6]Research Complex at Harwell, Harwell Science and Innovation Campus, Didcot, UK. [7]LCLS, SLAC National Accelerator Laboratory, Menlo Park, CA, USA [8]Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden.
*Corresponding author. Email: hugo.lebrette@univ-tlse3.fr (H.L.); jfkern@lbl.gov (J.K.); hogbom@dbb.su.se (M.H.)
†These authors contributed equally to this work.
‡Present address: Macromolecular Machines Laboratory, The Francis Crick Institute, London, UK.
§Present address: Sprint Bioscience, Huddinge, Sweden.

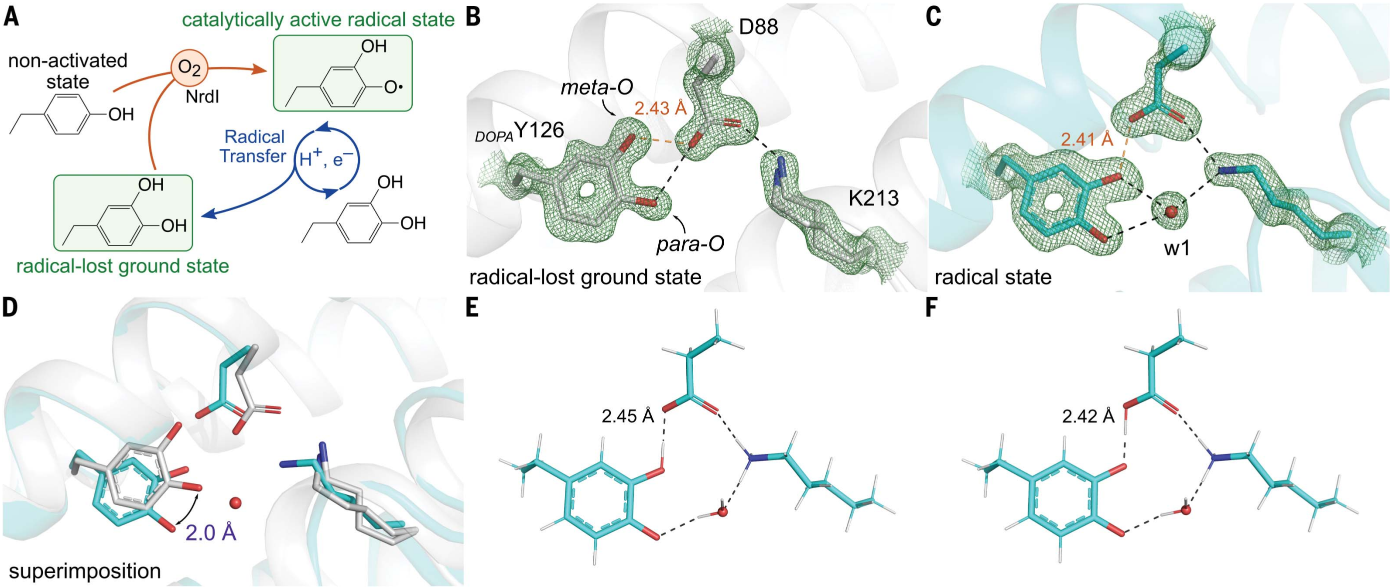


Fig. 1. Radical state and radical-lost ground state of *Mf*R2. (**A**) Outline of the proposed activation pathway of metal-free R2e, including the nonactivated state, catalytically active radical state, and radical-lost ground state (also known as "met" in canonical R2). The two states determined in this work are indicated in green. The structure of R2e in the nonactivated state has been determined previously (*13*, *14*). For clarity, chemical reactions are not strictly balanced. (**B**) Structure of *Mf*R2 in the radical-lost ground state obtained after chemical quenching by hydroxyurea (monomer A is shown). (**C**) Structure of *Mf*R2 in the radical state obtained from XFEL serial femtosecond crystallography showing a reorganization of the site compared with the ground state including a coupled movement of the $_{DOPA}$Y126-D88 dyad, the presence of a new water w1, and the inward conformation of K213. The short H-bond is highlighted in orange. (**D**) Superimposition of the ground and radical states. The 2-Å displacement of the DOPA *para*-O is marked in purple. Nitrogen and oxygen atoms are shown in blue and red, respectively. Carbons are shown in gray and cyan for the radical-lost ground state and radical state, respectively. Distance between atoms involved in H-bond interactions are in Å. Simulated annealing composite Omit $2F_o$–F_c electron density maps are shown in green and contoured at 2 σ. The structural changes are further illustrated in movie S1. (**E** and **F**) Using quantum mechanical calculations on the XFEL structure, the short H-bond between $_{DOPA}$Y126 and D88 can be reproduced by a DOPA radical state with the radical mainly located on the *para*-O and the proton located on *meta*-O (E) (neutral DOPA radical state) or D88 (F) (negatively charged DOPA radical state). For clarity, only a subset of residues included in the calculations is presented on the figure (see fig. S3 for full details).

calculated to 2.43 ± 0.04 Å, a length which could correspond to a low-barrier or single-well H-bond (*26*, *27*). In addition, $_{DOPA}$Y126 does not interact with any water, nor with K213 whose ε-ammonium group is facing away from $_{DOPA}$Y126 (Fig. 1B). Compared with previous R2e structures, differences and alternate conformations are observed throughout the structures, particularly variations in the conformation of $_{DOPA}$Y126, D88, and K213 (fig. S1A).

Structure of the radical state of *Mf*R2 by serial femtosecond crystallography

To circumvent photoreduction artifacts, we used XFEL serial femtosecond crystallography to determine the structure of the catalytically active radical state of *Mf*R2 at atomic resolution. The active radical-harboring protein was batch crystallized to produce a suspension of blue microcrystals used to collect room temperature serial femtosecond diffraction data at an XFEL source. The resulting dataset produces a model of the protein at 1.5-Å resolution, and the electron density map allows unambiguous interpretation (Fig. 1C). A short H-bond between $_{DOPA}$Y126 *meta*-O and D88 is present with a O–O distance calculated to 2.41 ± 0.05 Å using END/RAPID error analysis (*25*) (see Methods for details). A short H-bond is observed in both states of the protein and may play a special structural role as suggested in other cases (*28*), contributing to maintaining the integrity of the enzyme active site (*29*). The short H-bond may stabilize the interaction between $_{DOPA}$Y126 and D88 to ensure that no hydrogen is available to mediate a putative proton transfer from D88 to the $_{DOPA}$Y126 *para*-O, which would annul productive PCET. We note that this H-bonding structure results in a situation analogous to canonical R2 proteins in which a deprotonated aspartate is involved in metal coordination and not in proton transfer, thus forcing the latter to occur with a different nearby proton donor, suggested to be a metal-bound water molecule (*11*). Furthermore, it is tempting to speculate that this short H-bond is involved in redox tuning. By preventing the radical delocalization between the *meta*-O and *para*-O, the DOPA radical becomes electronically more similar to a tyrosyl radical, rather than a DOPA-semiquinone radical, in agreement with previous characterization of *Mf*R2 by electron paramagnetic resonance (EPR) (*13*).

A notable conformational shift takes place upon radical acquisition between $_{DOPA}$Y126 and D88: the dyad undergoes a coupled coplanar (but not coaxial) rotation of ~22°, with an additional rotation of the aromatic ring of ~12° along the C_β-C_1 axis. It results in a 2-Å displacement of the *para*-O carrying the main radical spin density away from D88 and a reorganization of the interaction pattern around $_{DOPA}$Y126 (Fig. 1D). The aspartate residue corresponding to D88 in *Mf*R2 is strictly conserved as the N-terminal metal-coordinating residue in Y•-harboring canonical R2 proteins and exhibits redox-induced conformational changes in metal-containing R2 proteins (*8*, *17*, *30*, *31*). Furthermore, coupled movements of the conserved radical-harboring Y have been observed to be redox-induced in R2b from *Bacillus anthracis* (*17*). For R2a from *E. coli* (*Ec*R2a), the conserved aspartate is proposed to form a H-bond with the reduced Y in the ground state. By contrast, single-crystal EPR experiments suggest that the Y• rotates away in the radical state, leading to a ~1-Å displacement of the radical-harboring oxygen and breaking the connection with the aspartate (*8*). A displacement of the Y• could also be

hypothesized from discrepancies observed between crystal structures and spectroscopic data for R2 proteins from *Bacillus anthracis* (*32*), *Salmonella typhimurium* (*33*), *Corynebacterium ammoniagenes* (*30*) and mouse (*19*). This type of rearrangement upon acquisition of the radical is principally similar to what we observe in *Mf*R2 structures and is less pronounced than movements proposed in other studies, which involve either translation of the main chain or larger Y• displacement by several Å (*20*, *21*, *24*).

In addition, a clearly defined water molecule (w1) mediates a new H-bond between $_{DOPA}$Y126 and the ε-ammonium group of K213, which displays a different orientation facing toward $_{DOPA}$Y126 (Fig. 1C). K213 adopts a single well-defined conformation different from previously published structures of active R2e determined using synchrotron radiation (*13*) (fig. S1B). The presence of a water molecule at a position similar to w1 has been observed previously in other RNR systems (*34*, *35*), and seems to be dependent on the redox state (*17*) (fig. S2). Moreover, superimposition with structures of R2 proteins shows that the new location of the K213 ε-ammonium group corresponds to the position of another water molecule that is metal-coordinated in canonical R2 proteins (fig. S2). This water is proposed to transfer a proton to Y• in the conformational gate initiating the PCET (*11*, *36*, *37*). Residue K213 was recently suggested by density functional theory to be a proton donor for radical transfer in R2e (*38*). Therefore, it may represent the water-equivalent proton donor in the case of R2e, and thus its conformational change could effectuate a comparable conformational gate.

Comparing the structures of the defined radical and radical-lost states determined here to previously solved structures of R2e proteins shows that no previous structure fully represents either the radical state or the radical-lost state (fig. S1). Although the proteins in prior work may have originally crystallized in the "active form," they appear to have suffered different degrees of x-ray–induced photoreduction during synchrotron data collection.

The XFEL structure can be reproduced in silico by a radical state

To evaluate whether the crystal structure obtained by XFEL femtosecond crystallography theoretically corresponds to a radical state, calculations were performed on the crystal structure active site. Based on quantum chemical geometry optimizations, the short interaction between $_{DOPA}$Y126 and D88 could be reproduced with the main radical character on the *para*-O. The proton could reside on either $_{DOPA}$Y126 or D88 as both states produced a short O–O distance (Fig. 1, E and F, and fig. S3, A and B). The calculated energy difference of 3 kcal/mol between the two states suggests that both states are accessible with slight favoring of the proton residing closer to $_{DOPA}$Y126•. In addition, various alternative $_{DOPA}$Y126 states were modeled by quantum mechanical calculations, none of them agreeing well with the experimental observations. In particular, a longer hydrogen bond with D88 is observed when $_{DOPA}$Y126 is modeled as a neutral DOPA, a DOPA quinone, or with the radical located on the *meta*-O (fig. S3, C to E). Furthermore, the calculated spin population of $_{DOPA}$Y126• in the protein active site models revealed an asymmetric distribution closer to a *meta*-substituted Y• than the fully delocalized character of an *ortho*-semiquinone, based on comparisons with calculated distributions in smaller models. This is fully consistent with spectroscopic data of the radical (*13*, *14*) and indicates a possible role for the short $_{DOPA}$Y126-D88 hydrogen bond to destabilize the otherwise potentially too stable semiquinone radical in the protein.

Molecular dynamics simulation starting from the XFEL structure of the radical state but with induced loss of the radical in silico, showed $_{DOPA}$Y126 movement with the dynamics dominated by the position of the radical-lost ground state, forming two H-bonds to D88 (fig. S3, F to H), consistent with the crystal structure. Altogether, our calculations support that the structure obtained by XFEL corresponds to the catalytically active radical state of the *Mf*R2 protein and are in agreement with previous EPR and UV–vis spectroscopic results (*13*), showing that the spin density distributes similar as in a *meta*-substituted tyrosyl radical rather than as in an *ortho*-semiquinone.

Specific protein rearrangements upon radical acquisition

The radical acquisition in *Mf*R2 leads to two major protein rearrangements that are of particular interest as they can be directly implicated in radical generation, stabilization, and transfer. The first major protein rearrangement takes place within the activating-oxidant path. The channel connecting the NrdI flavin cofactor to the R2b metal site (*39*, *40*), proposed as the $O_2^{\bullet-}$ route for activation, seems to be conserved in R2e (*14*). In the radical-lost ground state of *Mf*R2, in place of the metal center present in canonical R2, a continuous chain of well-defined H-bonded water molecules creates the link between the putative oxidant route and the $_{DOPA}$Y126-D88 dyad. By contrast, in the radical state of *Mf*R2, this water network is disrupted by Q91 which undergoes a large sidechain flip toward D88 (Fig. 2 and movie S1). In the radical-lost ground state, Q91 is involved in a H-bond network conserved in the R2b subclass (Q70 in R2b from *E. coli*), which lines the channel for oxidant transport to the Mn^{II}_2 active site (*40*). Our data suggest that Q91 could play a key role as it obstructs the putative oxidant channel in the radical state, preventing radical quenching by further oxidant access to the active site.

The second major protein rearrangement upon radical acquisition concerns the PCET route. In the immediate vicinity of the side chain of $_{DOPA}$Y126, residues L183 and F187 undergo large conformational changes in the radical

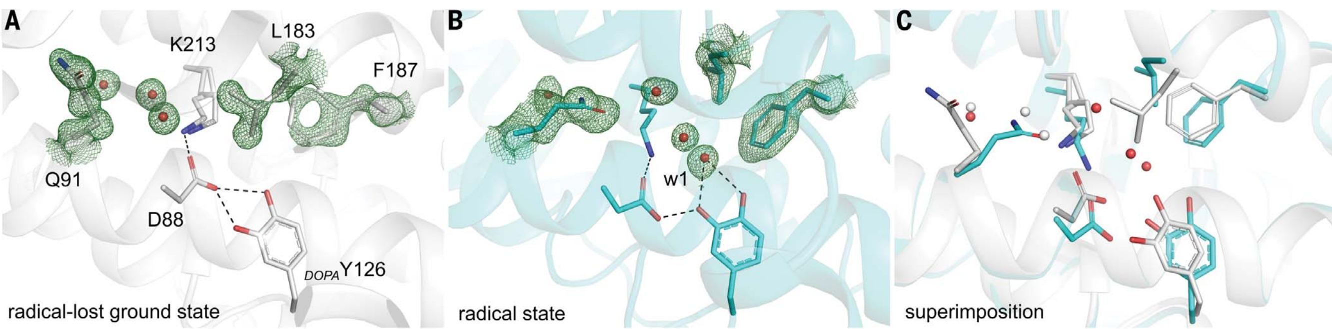


Fig. 2. Major conformational changes between the radical-lost ground state and the radical state. (**A**) Structure of *Mf*R2 in the radical-lost ground state obtained after chemical-quenching by hydroxyurea. (**B**) Structure of *Mf*R2 in the radical state obtained from XFEL serial femtosecond data showing reorganization of the site compared with the radical-lost ground state, including conformational changes of Q91, L183, and F187. In the radical state, Q91 displaces two water molecules, breaking the water H-bond network toward $_{DOPA}$Y126. Structural movements of L183 and F187 leave space for 3 water molecules interacting with D88, $_{DOPA}$Y126, and K213. Simulated annealing composite Omit $2F_o–F_c$ electron density maps are shown in green and contoured at 1.5 σ. (**C**) Superimposition of the structures of *Mf*R2 in the radical state (cyan) and radical-lost ground state (gray).

state of *Mf*R2, leaving space for the positioning of three water molecules (including w1) and creating a water-mediated H-bond network between $_{DOPA}$Y126, D88, K213, and Q91 (Fig. 2 and movie S1). Residues L183 and F187 belong to helix αE which forms a distorted π-helix conserved across many ferritin superfamily members. The π-helix conformation of helix αE is believed to play a functional role (*41*, *42*), and is known to undergo redox-induced structural rearrangements (*17*, *30*, *34*). Residues L183 and F187 are conserved in all R2 proteins, the position of L183 being either F or L (in 91 and 9% of sequences, respectively) and F187 being conserved in 99% of the sequences. In none of the R2 crystal structures solved to date do these residues exhibit conformations resembling those in the *Mf*R2 radical state (fig. S4). In the first solved structure of R2, it was noted that the radical-harboring tyrosyl oxygen was surrounded by a conserved hdrophobic pocket formed by residues F208, F212, and I234 in *Ec*R2a (equivalents to L183, F187, and I209 in *Mf*R2, respectively) (*43*). The major function of these residues was proposed through mutational studies to contribute to the tyrosyl radical stability by insulating the radical-harboring tyrosyl oxygen (*44*). Our observation of hitherto unseen movements of these radical-shielding residues further implicates them in radical control and suggests their involvement in gating of the ribonucleotide reductase PCET mechanism. The specific local rearrangements at the radical site also translate to movements of the protein backbone and global structural changes of the protein scaffold protruding to the R1 interaction surface and the radical transfer path (movie S2). It has previously been shown that the active R1-R2 complex exhibits tighter binding after radical initiation (*45*–*47*). We propose that these global structural changes observed in R2 provide a mechanism by which the R1-R2 binding properties can be modulated during the catalytic cycle.

Model of conformational gating for radical transfer initiation

In *Mf*R2, the catalytically active radical state and the radical-lost ground state are interconvertible by quenching the radical and through NrdI-mediated reoxidation of $_{DOPA}$Y126 (Fig. 1A). Based on our structural data, we propose a model of the conformational gating orchestrated by R2 after radical acquisition, which is a prelude to the radical translocation to the R1 subunit (Fig. 3). This model proceeds in three steps. First, the oxidation of $_{DOPA}$Y126 leads to a repulsion between its *para*-O and D88 due to the removal of the H-bonding hydrogen atom, resulting in the 2-Å displacement of the $_{DOPA}$Y126 *para*-O (Fig. 3A). Secondly, this triggers a cascade of structural changes to shield the radical and prepare its transfer: Q91 blocks the access to further oxidant, L183 and F187 reshape the insulating pocket around the radical, and water w1 connects the $_{DOPA}$Y126 radical-carrying oxygen with the PCET route (Fig. 3B). This also leads to global structural rearrangements of protein R2, including the R2-R1 interaction surface and binding of protein R1 (Fig. 3C and movie S2). Thirdly, the formation of the R1-R2 complex results in the ordering of the full R2 C-terminal tail at the R1-R2 interface [as demonstrated in (*7*)], completing the electron transfer path and inducing the injection of an electron to reduce $_{DOPA}$Y126•, e.g., through the conserved W52 and/or Y325 (corresponding to W48 and Y356 in *Ec*R2a), as previously suggested (*12*) (Fig. 3D). Coupled to this event, a proton transfer from K213 to $_{DOPA}$Y126• occurs by means of the water w1 [as proposed in (*38*)] and initiates the long-range radical translocation. Figure 3 summarizes the proposed steps of conformational gating in R2e. The interaction with R1 is modeled based on a superposition of the cryo-EM structure of the R1-R2 holocomplex from *E. coli* (*7*).

This conformational gating model for initiating the radical translocation could be common to class I RNR systems. In PCET, proton

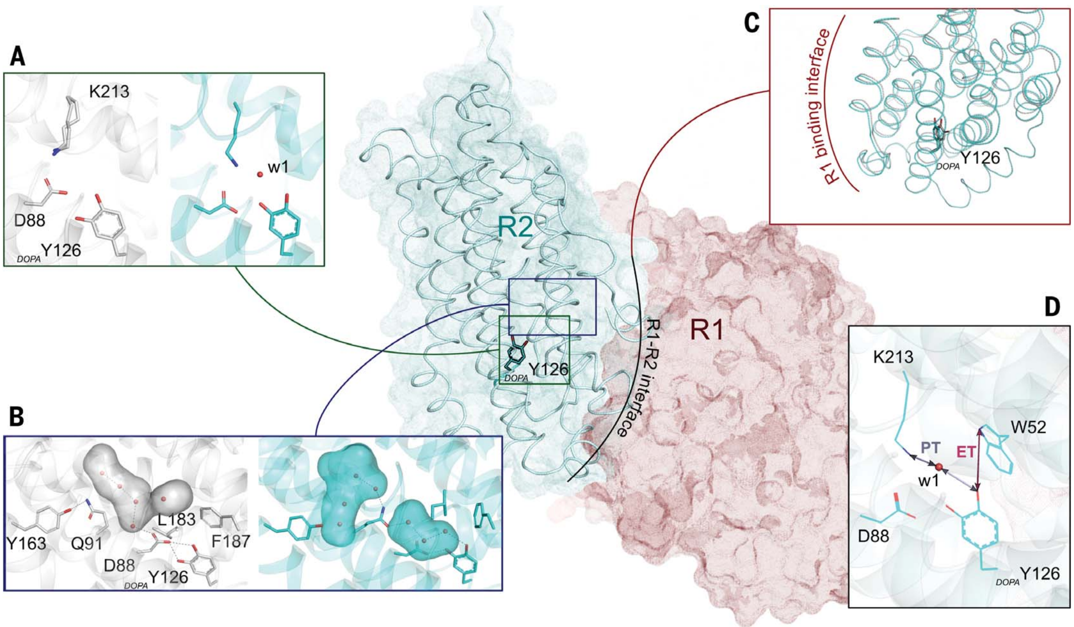


Fig. 3. Model of conformational gating and radical transfer during PCET in R2e. (**A**) Oxidation of the radical-lost ground state $_{DOPA}$Y126 (gray) displaces $_{DOPA}$Y126-D88 and introduces a water molecule (w1) to facilitate the proton transfer from the redirected amino group of K213 (cyan). (**B**) Concurrently, the cavity surrounding the generated $_{DOPA}$Y126 radical is reshaped, shown in surface representation (radical-lost state and radical state in gray and cyan, respectively). Primarily, Q91 flips away from Y163 to block access of any further oxidant from NrdI through the R2-NrdI channel and L183 and F187 flip to reshape the cavity and facilitate electron transfer from W52 to $_{DOPA}$Y126. (**C**) Structure superposition of *Mf*R2 in the radical (cyan) and radical-lost (gray) states at the R1 binding interface. Reshaping of the pockets surrounding the $_{DOPA}$Y126 leads to conformational changes protruding to the R2-R1 interaction surface and binding to protein R1 completes the PCET pathway. (**D**) PCET is initiated by a proton transfer (PT) from the amino group of K213 to $_{DOPA}$Y126 mediated by w1, and simultaneously an electron transfer from W52 to $_{DOPA}$Y126, thus translocating the radical toward the active site C394 of R1. R1 (represented in red) is modeled in complex with the radical-state *Mf*R2 using the cryo-EM structure of the R1-R2 holocomplex from *E. coli* (PDB ID: 6w4x) (*7*).

transfer occurs through H-bond networks and requires the proton donor and acceptor to be within a standard ~2.8-Å H-bond distance (*48*). As a displacement of the radical-harboring Y upon radical acquisition is also observed by spectroscopy in other R2 proteins (*8*, *19*, *30*, *32*, *33*), an intermediary may be required between the metal-bound water and Y•. In *Mf*R2 the additional water w1 gained upon radical acquisition represents the missing piece of the puzzle, connecting the radical to K213 (Fig. 3D). We note that an analogous binding site for water has been observed in other R2 proteins (fig. S2), and that the ε-ammonium group of K213 is located at the position of the metal-bound water in metal-containing R2s (fig. S2). Therefore, we speculate that, similarly as in R2e, a water in this position links the radical to the metal-bound water proposed to be the proton donor in canonical R2 proteins and gates the first proton transfer to initiate radical translocation to R1.

REFERENCES AND NOTES

1. M. Högbom, B.-M. Sjöberg, G. Berggren, in *eLS* (Wiley, 2020), vol. 1, pp. 375–393.
2. J. Stubbe, D. G. Nocera, *J. Am. Chem. Soc.* **143**, 13463–13472 (2021).
3. A. Ehrenberg, P. Reichard, *J. Biol. Chem.* **247**, 3485–3488 (1972).
4. L. Miret-Casals *et al.*, *ACS Omega* **3**, 17057–17069 (2018).
5. Y. Xie *et al.*, *J. Biomed. Sci.* **29**, 32 (2022).
6. R. Banerjee, V. Srinivas, H. Lebrette, in *Macromolecular Protein Complexes IV*, J. R. Harris, J. Marles-Wright, Eds. (Springer, 2022), vol. 99 of *Subcellular Biochemistry*, pp. 109–153.
7. G. Kang, A. T. Taguchi, J. Stubbe, C. L. Drennan, *Science* **368**, 424–427 (2020).
8. M. Högbom *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3209–3214 (2003).
9. A. R. Offenbacher, I. R. Vassiliev, M. R. Seyedsayamdost, J. Stubbe, B. A. Barry, *J. Am. Chem. Soc.* **131**, 7496–7497 (2009).
10. K. Yokoyama, U. Uhlin, J. Stubbe, *J. Am. Chem. Soc.* **132**, 15368–15379 (2010).
11. B. Wörsdörfer *et al.*, *J. Am. Chem. Soc.* **135**, 8585–8593 (2013).
12. S. Y. Reece, M. R. Seyedsayamdost, *Essays Biochem.* **61**, 281–292 (2017).
13. V. Srinivas *et al.*, *Nature* **563**, 416–420 (2018).
14. E. J. Blaesi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 10022–10027 (2018).
15. J. A. Cotruvo Jr., J. Stubbe, *Metallomics* **4**, 1020–1036 (2012).
16. E. F. Garman, *Acta Crystallogr. D* **66**, 339–351 (2010).
17. K. Grāve *et al.*, *J. Biol. Inorg. Chem.* **24**, 849–861 (2019).
18. K. G. V. Sigfridsson *et al.*, *J. Biol. Chem.* **288**, 9648–9661 (2013).
19. K. R. Strand *et al.*, *J. Biol. Chem.* **279**, 46794–46801 (2004).
20. B. A. Barry *et al.*, *J. Phys. Chem. Lett.* **3**, 543–554 (2012).
21. A. R. Offenbacher, L. A. Burns, C. D. Sherrill, B. A. Barry, *J. Phys. Chem. B* **117**, 8457–8468 (2013).
22. H. N. Chapman *et al.*, *Nature* **470**, 73–77 (2011).
23. N. Saban, M. Bujak, *Cancer Chemother. Pharmacol.* **64**, 213–221 (2009).
24. A. R. Offenbacher, B. A. Barry, *J. Phys. Chem. B* **124**, 345–354 (2020).
25. P. T. Lang, J. M. Holton, J. S. Fraser, T. Alber, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 237–242 (2014).
26. C. L. Perrin, J. B. Nielson, *Annu. Rev. Phys. Chem.* **48**, 511–544 (1997).
27. C. L. Perrin, *Acc. Chem. Res.* **43**, 1550–1557 (2010).
28. D. A. Nichols *et al.*, *J. Am. Chem. Soc.* **137**, 8086–8095 (2015).
29. M. T. Kemp, E. M. Lewandowski, Y. Chen, *Biochim. Biophys. Acta. Proteins Proteomics* **1869**, 140557 (2021).
30. N. Cox *et al.*, *J. Am. Chem. Soc.* **132**, 11197–11213 (2010).
31. D. T. Logan *et al.*, *Structure* **4**, 1053–1064 (1996).
32. E. Torrents, M. Sahlin, D. Biglino, A. Gräslund, B.-M. Sjöberg, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 17946–17951 (2005).
33. M. Galander, M. Uppsten, U. Uhlin, F. Lendzian, *J. Biol. Chem.* **281**, 31743–31752 (2006).
34. M. Eriksson, A. Jordan, H. Eklund, *Biochemistry* **37**, 13359–13369 (1998).
35. M. Högbom, Y. Huque, B.-M. Sjöberg, P. Nordlund, *Biochemistry* **41**, 1381–1389 (2002).
36. P. E. M. Siegbahn, L. Eriksson, F. Himo, M. Pavlov, *J. Phys. Chem. B* **102**, 10622–10629 (1998).
37. J. Stubbe, D. G. Nocera, C. S. Yee, M. C. Y. Chang, *Chem. Rev.* **103**, 2167–2201 (2003).
38. J. Zou, Y. Chen, W. Feng, *Comput. Struct. Biotechnol. J.* **20**, 1111–1131 (2022).
39. J. John *et al.*, *eLife* **11**, e79226 (2022).
40. A. K. Boal, J. A. Cotruvo Jr., J. Stubbe, A. C. Rosenzweig, *Science* **329**, 1526–1530 (2010).
41. D. Lundin, A. M. Poole, B.-M. Sjöberg, M. Högbom, *J. Biol. Chem.* **287**, 20565–20575 (2012).
42. R. B. Cooley, D. J. Arp, P. A. Karplus, *J. Mol. Biol.* **404**, 232–246 (2010).
43. P. Nordlund, B. M. Sjöberg, H. Eklund, *Nature* **345**, 593–598 (1990).
44. M. Ormö *et al.*, *J. Biol. Chem.* **270**, 6570–6576 (1995).
45. B. L. Greene, D. G. Nocera, J. Stubbe, *J. Biol. Chem.* **293**, 10413–10414 (2018).
46. E. C. Minnihan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 3835–3840 (2013).
47. B. L. Greene *et al.*, *Annu. Rev. Biochem.* **89**, 45–75 (2020).
48. A. Migliore, N. F. Polizzi, M. J. Therien, D. N. Beratan, *Chem. Rev.* **114**, 3381–3465 (2014).
49. H. Lebrette, K. Roos, V. Kaila, M. Högbom, QM/MM and MD result files for the MfR2e system, SciLifeLab Data Repository (2023).

ACKNOWLEDGMENTS

We thank R. Bolotovsky and I. D. Young for help with data processing and the staff at LCLS. We acknowledge the Paul Scherrer Institute, Villigen, Switzerland for provision of synchrotron radiation beamtime (proposals 20161653 and 20182304), and the staff of beamline X06SA (PXI) at the Swiss Light Source for assistance. We acknowledge SOLEIL for provision of synchrotron radiation facilities (proposal 20180270) and we thank the staff of beamlines PROXIMA 1 and PROXIMA 2A for assistance. Research was carried out at the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory (proposal LU50), supported by the DOE Office of Science, OBES under contract DE-AC02-76SF00515. XFEL data processing was performed in part at the National Energy Research Scientific Computing Center, supported by the DOE Office of Science, contract DEAC02-05CH11231. The Rayonix detector used at LCLS was supported by the NIH grant S10 OD023453. Experiments at the LCLS were supported by the NIH grant P41GM139687. D.L. acknowledges the Centre for Ecology and Evolution in Microbial model Systems – EEMiS, Linnaeus University. A.S. and V.R.I.K. acknowledge the National Academic Infrastructure for Supercomputing in Sweden (NAISS 2023/1-31) for computational resources. **Funding:** The authors acknowledge financial support of this work by the Swedish Research Council (2021-03992 to M.H. and 2019-01400 to B.-M.S.), by the European Research Council (HIGH-GEAR 724394 to M.H.), by the Knut and Alice Wallenberg Foundation (2017.0275 and 2019.0436 to M.H., 2019.0251 to V.R.I.K.), by the Swedish Cancer Foundation (20 1210 PjF to B.-M.S.), by the Swedish Foundation for Strategic Research (ICA16-0040 to K.R.), by the Swedish National Infrastructure for Computing (SNIC 2021/5-137 to K.R.), by the Director, Office of Science, Office of Basic Energy Sciences (OBES), Division of Chemical Sciences, Geosciences, and Biosciences (CSGB), Department of Energy (DOE) (to J.Y., V.K.Y., and J.K.), by the National Institutes of Health (NIH) grants GM133081 (to K.D.S.), GM055302 and 1R35GM149528-01 (to V.K.Y.), GM110501 (to J.Y.) GM126289 (to J.K.), GM117126 (to N.K.S.), by Diamond Light Source, the UK Science and Technology Facilities Council (STFC), and Biotechnology and Biological Sciences Research Council (to A.M.O., P.A., and A.Bu.), by a Wellcome Investigator Award (210734/Z/18/Z to A.M.O.), by a Royal Society Wolfson Fellowship (RSWF\R2\182017 to A.M.O.), and by an EMBO Long-Term Fellowship (ALTF 952-2022 to A.S.). **Author contributions:** H.L. performed crystallography, processed and analyzed synchrotron and XFEL data, and wrote the manuscript; V.S. performed protein production and crystallography; M.H. conceived and led the study; K.R., A.S., and V.R.I.K. performed computational studies; H.L., V.S., J.J., R.K., J.K., K.R., and M.H. edited the manuscript; D.L. performed bioinformatics; B.-M.S. performed study design; C.P., I.-S.K., S.G., A.M.O., F.D.F., and J.K. developed, tested and ran the XFEL sample delivery system; H.L., A.S.B., K.D.S., A.Bh., A.Bu., and N.K.S. processed and analyzed XFEL data; F.D.F., R.A.-M., and A.Ba. operated the MFX instrument; H.L., J.J., O.A., C.P., I.-S.K., A.S.B., S.G., K.D.S., A.Bh., P.S.S., A.Bu., P.A., A.M.O., F.D.F., A.Ba., N.K.S., V.K.Y., J.Y., J.K., and M.H. performed the XFEL experiment at LCLS. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** Atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) with the following codes: catalytically active radical state solved by XFEL, 8bt3; radical-lost ground state, 8bt4. *In silico* models and output are available at the SciLifeLab Data Repository (*49*).

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh8160
Materials and Methods
Figs. S1 to S5
Table S1
References (*50*–*66*)
MDAR Reproducibility Checklist
Movies S1 and S2

Submitted 17 March 2023; accepted 30 August 2023
10.1126/science.adh8160

WORKING LIFE

By **Daniela Osorio Rodriguez**

A healthier routine

"Are you all right?" the student asked as she passed me in the hallway. I was in the final stretch of lab work before writing my dissertation. The student, who was 4 years behind me in the same Ph.D. program, went on to say, "I see people from your cohort working so much, putting in extra hours, even at night." She was concerned about me, but afterward, I realized she also wondered whether the same was expected of her. She didn't know I had quietly rethought my work routine midway through my Ph.D. In that moment I knew I had a responsibility to tell students who followed in my footsteps that there was another, more sustainable way to approach their work. It's a message I wish I'd heard when I began graduate school.

As a first-year Ph.D. student, I was lucky to have kind and understanding advisers. But I felt overwhelming pressure to become a flawless student. I kept finding a million reasons why I was coming into graduate school at a disadvantage compared with my peers. As a student from Colombia, I had been educated at institutions my professors and colleagues had never heard of, I spoke with an accent that was hard to understand, I blushed every time I spoke in public, and I was older than most of the people in my cohort. I wanted to be the best possible representation of an international Latina student. But I couldn't shake the feeling that I was an impostor in my new academic setting.



"I'd fallen into the trap of assuming my colleagues' work habits were the standard for belonging."

The lack of firm deadlines and clear expectations in my program added to my anxieties and left me without a clear path forward. So, I instead looked around to get an idea of how I could fit in and impress my colleagues. I noticed that most of the people around me seemed to be working all the time. I began to spend many weekday nights and weekends in the lab, even when it wasn't strictly required for my project.

After I passed my qualifying exam at the end of the first year, though, my strategy backfired. I crashed from burnout. The fatigue took a toll on my productivity and motivation. During one of many mindless Sundays in the office, I realized I'd fallen into the trap of assuming my colleagues' work habits were the standard for belonging and being worthy of respect—and it wasn't sustainable.

From then on, I stopped counting the number of work hours I was putting in, leaving space every day for exercising and doing activities that I enjoyed. My new approach brought guilt and fearful feelings at first, but I pushed through with the aid of meditation and journaling. Eventually, I realized nobody was keeping track of the amount of time I spent in the lab anyway and it wasn't something I needed to worry about.

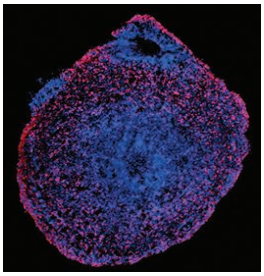
I noticed that having more time for myself made me more rested and motivated when I went back to work. And as I developed a more positive mindset, I became aware of the unique contributions I could make to my research precisely because of my distinctive background. My impostor syndrome began to dissolve, and my confidence grew.

I continued quietly along this path for years, until that conversation in the hallway. The student was making the same mistake I did at her stage, setting for herself the unhealthy standard of what the older students were doing. I realized then that I could be a positive role model for others, and I had to be more vocal about what had worked well for me.

So, when I ran into her again, weeks later, I told her that even though I was feeling a lot of pressure in the final year of my Ph.D., I was putting my mental and physical health first and wasn't working overtime. I also began to talk with other students about how quality sleep, physical activity, healthy eating, and reasonable work hours were a daily part of my routine.

The last year of my Ph.D. was hard. But I did my best to stick to a healthy work-life balance, and I felt good knowing I was doing my part to set a good example for my peers and combat the workaholic culture around me. It's an approach I plan to continue throughout my career. ■

Daniela Osorio Rodriguez is a postdoc at the California Institute of Technology. Send your career story to SciCareerEditor@aaas.org.



Spinning Disk Microscope System

CrestOptics has announced the launch of a new spinning disk confocal microscope system called CICERO. Focusing on essential components of epi-fluorescence and confocal microscopy, CICERO was first developed to expand the company's portfolio and make high-end imaging more accessible to the scientific community at large. The result is a compact and cost-effective imaging system, comprised of a spinning disk confocal instrument that can be easily integrated into existing workflows with maximum configuration flexibility to support a seamless and user-friendly transition.

CrestOptics

For info: +39 066-166-0508

crestoptics.com/cicero/

Versatile Flow Cytometry Dye

Life science and clinical diagnostic supplier Bio-Rad Laboratories, Inc. has announced the release of its latest StarBright Dye range, StarBright Red 670. The 29th in the StarBright product line, Red 670 provides researchers with validated flow antibodies against key immunology targets, conjugated to proprietary fluorescent nanoparticles. Serving as an expansion of its current range of highly validated antibodies, Red 670 is conjugated to StarBright Blue 700 and StarBright Violet 610 Dyes. Compared with former StarBright Dyes, Red 670 offers additional human and mouse targets. It also introduces dog, cow, and pig targets—a first for StarBright—a new range providing for greater freedom and flexibility. In flow cytometry, the additional targets work both with conventional panels and with full-spectrum multicolor panels, expanding research capabilities across veterinary immunology research. Like all StarBright Dyes, Red 670 needs no special buffers for compatibility with the Bio-Rad ZE5 Cell Analyzer and S3e Cell Sorter, most flow cytometers, and most experimental protocols. Minimal lot-to-lot variation ensures reproducible and consistent staining, and the dyes are resistant to photobleaching with no loss of signal in fixation.

"Now totaling 29 dyes for the 355, 405, 488, 561, and 640 nm lasers, the Bio-Rad StarBright Dye range is currently the largest series of dyes for conventional and full spectrum flow cytometry," says Dr. Mike Blundell, Flow Cytometry Product Manager at Bio-Rad. "By extending our existing portfolio of 35 antibody markers to include another 29 human, mouse, and now cow, dog, and pig targets, we can offer our customers greater flexibility and choice when designing multicolor flow cytometry panels, including for translational and immunological research in veterinary species."

Bio-Rad

For info: +1 510-356-7909

www.bio-rad.com/en-us/

Dual-Cap Microcentrifuge Tube

Azenta Life Sciences has released Cap2, a new 0.2ml PCR microcentrifuge tube featuring a dual-cap design, to preserve the integrity of samples used in genetic testing. Acting as both a collection and processing tube, Cap2 can minimize sample transfer and improve efficiency by reducing wastage. Designed to ensure sample traceability in fields like pre-implantation genetic testing, each Cap2 tube has both a unique 2D datamatrix code and a human-readable numerical identifier to prevent sample tracking errors. Like most modern tubes, the new Cap2 was designed for broad compatibility. It works in conjunction with automated systems for code reading, capping, and sample management—from Azenta and all other industry-recognized manufacturers. Manufactured from high-quality virgin polypropylene in a Class 8 clean room, each batch of tubes is tested and certified as free from endotoxins, DNase/RNase, leachables, and extractables.

Azenta

For info: +1 888-229-3682

www.azenta.com/products/0.2ml-dual-cap-sample-collection-pcr-tube

Fixed-focus Radioactive Process Lens

Resolve Optics has announced the availability of lenses used to monitor radioactive processes in nuclear power plants and nuclear waste reprocessing facilities. The fixed-focus lenses come in the standard range of 6mm, 9mm, and 25mm focal length. Designed to provide a focal length of 6mm or 9mm respectively, the Model 214 and 286 radiation-tolerant lenses deliver high-quality images with minimal geometric distortion from 400-750nm. The lenses are designed specifically for use with a 2/3-inch image format radiation-resistant tube and CMOS cameras. Key features of Models 214 and 286 include their wide field of view and variable iris control. For producing sharp images without a strong yellow tint, Resolve Optics developed the Model 313 25mm focal length fixed-focus lens. Although yellow-tinting has traditionally been among the limitations of using radiation-tolerant lenses on color CMOS cameras, the cerium-doped glass in the Model 313 produces high-quality natural color images.

Resolve Optics

For info: +44-1494-777100

www.resolveoptics.com/radiation-resistant-lenses-2/

Super-Resolution Confocal Microscope System

Nikon recently released the AX / AX R with NSPARC confocal microscope system, which combines Nikon's flagship AX / AX R confocal system with the Nikon SPatial ARray Confocal (NSPARC) detector to realize next-level sensitivity and super-resolution with a confocal instrument. The NSPARC is based on a single-pixel photon counter (SPPC) array detector that brings the power of image scanning microscopy (ISM), an innovative technique capable of providing 100 nm lateral (XY) resolution, to an otherwise traditional point-scanning confocal system architecture. Not only does the NSPARC detector support super-resolution imaging, but it has superior sensitivity compared to traditional confocal detectors due to its ultra-low noise profile. This allows users to make full and practical use of the increased resolving power, even in challenging experimental conditions. Combine NSPARC detection with high-speed resonant scanning for a robust live-cell compatible video-rate super-resolution imaging solution.

Nikon Instruments

For info: +1-631-547-8500

www.microscope.healthcare.nikon.com/ax

Electronically submit your new product description or product literature information! Go to www.science.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

DOWNLOAD MORE MAGAZINES AT HTTPS://FREEMAGAZINES.TOP